

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2001, 01:20:46 ; Search time 1967.09 Seconds
(without alignment)
12888.354 Million cell updates/sec

Title: US-09-707-121-1
Perfect score: 2682
Sequence: 1 algtlcccccaaaagaaacg.....lccacalcclcaaatataga 2682

Scoring table: IDENTIFY NDC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 47646750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*

1: qb_est1: *
2: qb_est2: *
3: qb_est3: *
4: qb_est4: *
5: qb_est5: *
6: qb_est6: *
7: qb_est7: *
8: qb_est8: *
9: qb_est9: *
10: qb_est10: *
11: qb_est11: *
12: qb_est12: *
13: qb_est13: *
14: qb_est14: *
15: qb_est15: *
16: qb_est16: *
17: qb_est17: *
18: qb_est18: *
19: qb_est19: *
20: qb_est20: *
21: qb_est21: *
22: qb_est22: *
23: qb_est23: *
24: qb_est24: *
25: qb_est25: *
26: qb_est26: *
27: qb_est27: *
28: qb_est28: *
29: qb_est29: *
30: qb_est30: *
31: qb_est31: *
32: qb_est32: *
33: qb_est33: *
34: qb_est34: *
35: qb_est35: *
36: qb_est36: *
37: qb_est37: *
38: qb_est38: *
39: qb_est39: *
40: qb_est40: *
41: qb_est41: *
42: qb_est42: *
43: qb_est43: *

44: em_esthum10: *
45: em_esthum11: *
46: em_esthum12: *
47: em_esthum13: *
48: em_esthum14: *
49: em_esthum15: *
50: em_esthum16: *
51: em_esthum17: *
52: em_esthum18: *
53: em_esthum19: *
54: em_esthum20: *
55: em_esthum21: *
56: em_esthum22: *
57: em_esthum23: *
58: em_esthum24: *
59: em_esthum25: *
60: em_esthum26: *
61: em_esthum27: *
62: em_esthum28: *
63: em_esthum29: *
64: em_esthum30: *
65: em_esthum31: *
66: em_esthum32: *
67: em_esthum33: *
68: em_esthum34: *
69: em_esthum35: *
70: em_esthum36: *
71: em_esthum37: *
72: em_esthum38: *
73: em_esthum39: *
74: em_esthum40: *
75: em_esthum41: *
76: em_esthum42: *
77: em_esthum43: *
78: em_esthum44: *
79: em_esthum45: *
80: em_esthum46: *
81: em_esthum47: *
82: em_esthum48: *
83: em_esthum49: *
84: em_esthum50: *
85: em_esthum51: *
86: em_esthum52: *
87: em_esthum53: *
88: em_esthum54: *
89: em_esthum55: *
90: em_esthum56: *
91: em_esthum57: *
92: em_esthum58: *
93: em_esthum59: *
94: em_esthum60: *
95: em_esthum61: *
96: em_esthum62: *
97: em_esthum63: *
98: em_esthum64: *
99: em_esthum65: *
100: em_esthum66: *
101: em_esthum67: *
102: em_esthum68: *
103: em_esthum69: *
104: em_esthum70: *
105: em_esthum71: *
106: em_esthum72: *
107: em_esthum73: *
108: em_esthum74: *
109: em_esthum75: *
110: em_esthum76: *
111: em_esthum77: *
112: em_esthum78: *
113: em_esthum79: *
114: em_esthum80: *
115: em_esthum81: *
116: em_esthum82: *

```

117: qb_cst48:*
118: qb_cst49:*
119: qb_cst50:*
120: qb_cst51:*
121: qb_cst52:*
122: qb_cst53:*
123: qb_cst54:*
124: qb_cst55:*
125: qb_cst56:*
126: qb_cst57:*
127: qb_cst58:*
128: qb_cst59:*
129: qb_cst60:*
130: qb_cst61:*
131: qb_cst62:*
132: qb_cst63:*
133: qb_cst64:*
134: qb_cst65:*
135: qb_cst66:*
136: qb_cst67:*
137: qb_cst68:*
138: qb_cst69:*
139: qb_cst70:*
140: qb_cst71:*
141: qb_cst72:*
142: qb_cst73:*
143: qb_cst74:*
144: qb_cst75:*
145: qb_cst76:*
146: qb_cst77:*
147: qb_cst78:*
148: qb_cst79:*
149: qb_cst80:*
150: qb_cst81:*
151: qb_cst82:*
152: qb_cst83:*
153: qb_cst84:*
154: qb_cst85:*
155: qb_cst86:*
156: qb_cst87:*
157: qb_cst88:*
158: qb_cst89:*
159: qb_cst90:*
160: qb_cst91:*
161: qb_cst92:*
162: qb_cst93:*
163: qb_cst94:*
164: qb_cst95:*
165: qb_cst96:*
166: qb_cst97:*
167: qb_cst98:*
168: qb_cst99:*
169: qb_cst100:*
170: qb_cst101:*
171: qb_cst102:*
172: qb_cst103:*
173: qb_cst104:*
174: qb_cst105:*
175: qb_cst106:*
176: qb_cst107:*
177: qb_cst108:*
178: qb_cst109:*
179: qb_cst110:*
180: qb_cst111:*
181: qb_cst112:*
182: qb_cst113:*
183: qb_cst114:*
184: qb_cst115:*
185: qb_cst116:*
186: qb_cst117:*
187: qb_cst118:*
188: qb_cst119:*
189: qb_cst120:*

```

```

190: qb_cst110:*
191: qb_cst111:*
192: qb_cst112:*
193: qb_cst113:*
194: qb_cst114:*
195: qb_cst115:*
196: qb_cst116:*
197: qb_cst117:*
198: qb_cst118:*
199: qb_cst119:*
200: qb_cst120:*
201: qb_cst121:*
202: qb_cst122:*
203: qb_cst123:*
204: qb_cst124:*
205: qb_cst125:*
206: qb_cst126:*
207: qb_cst127:*
208: qb_cst128:*
209: qb_cst129:*
210: qb_cst130:*
211: qb_cst131:*
212: qb_cst132:*
213: qb_cst133:*
214: qb_cst134:*
215: qb_cst135:*
216: qb_cst136:*
217: qb_cst137:*
218: qb_cst138:*
219: qb_cst139:*
220: qb_cst140:*
221: qb_cst141:*
222: qb_cst142:*
223: qb_cst143:*
224: qb_cst144:*
225: qb_cst145:*
226: qb_cst146:*
227: qb_cst147:*
228: qb_cst148:*
229: qb_cst149:*
230: qb_cst150:*
231: qb_cst151:*
232: qb_cst152:*
233: qb_cst153:*
234: qb_cst154:*
235: qb_cst155:*
236: qb_cst156:*
237: qb_cst157:*
238: qb_cst158:*
239: qb_cst159:*
240: qb_cst160:*
241: qb_cst161:*
242: qb_cst162:*
243: qb_cst163:*
244: qb_cst164:*
245: qb_cst165:*
246: qb_cst166:*
247: qb_cst167:*
248: qb_cst168:*
249: qb_cst169:*
250: qb_cst170:*
251: qb_cst171:*
252: qb_cst172:*
253: qb_cst173:*
254: qb_cst174:*
255: qb_cst175:*
256: qb_cst176:*
257: qb_cst177:*
258: qb_cst178:*

```

pred. No. is the number of results predicted by change to have a score greater than or equal to the score of the result point printed, and is derived by analysts of the total score distribution.


```

RESULTS 10
LOCUS      H16878      453 bp      mRNA      EST      29-JUN-1995
DEFINITION ym34f07 r1 Soares infant brain INH Homo sapiens cDNA clone
            IMAGE:50126 5' similar to Sp-B0393.2 CE00855 ; mRNA sequence.
VERSION    H16878
XREF       H16878.1 GI:883118
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 453)
            Hillier, L., Clark, N., Duboue, T., Elliston, K., Hawkins, M., Holman
            M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
            Rifkin, B., Rohlfing, J., Soares, M., Tan, F., Trevisakis, E., Waterston
            R., Williamson, A., Wohldmann, P. and Wilson, R.
            The WashU-Merck EST Project
            Unpublished (1995)
            Contact: Wilson, RK
            Washington University School of Medicine
            444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1740
            High quality sequence stops: 430
            Source: IMAGE Consortium, LHM.
            This clone is available royalty-free through LHM ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1740 Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 430.
FEATURES
SOURCE
            location/Qualifiers
            1..453
            /organism="Homo sapiens"
            /db_xref="GDB:422936"
            /db_xref="taxon:9606"
            /clone="IMAGE:50126"
            /clone_11b="Soares infant brain INH"
            /sex="female"
            /dev_stage="73 days post natal"
            /lab_host="DH10B (ampicillin resistant)"
            /note="Organ: whole brain; Vector: Lambda BA; Site:1: Not
            1; Site:2: Hind III; 1st strand cDNA was primed with a NOT
            1; -01190(47) primer 15'
            AACTGGAACATCCGCGCCGACGAGATTTTCTTTTCTTTT 3';
            double-stranded cDNA was ligated to Hind III adaptors
            (pharmacia), digested with Not I and directionally cloned
            into the Not I and Hind III sites of the Lambda BA vector.
            Library went through one round of normalization. Library
            constructed by Bento Soares and M. Patricia Ronaldo."
BASE COUNT      108 a      100 c      81 g      162 t      2 others
ORIGIN
Query Match      13.8%      Score 170; DB 157; Length 453;
Post local Similarity 99.7%      Pred. No. 7.1e-92;
Matches 381; Conservative 0; Mismatches 0; Indels 1; Gaps 1.
QY 1595 laaagcctgaagatgctcaacgcagctgctgctgctatgcaagatcctgactcattt 1654
      1 TAAAGCCTGGGTAGCTCTCATCTCGATCTTGCTGATTTGGCAAGCTCTTGACTCAGTTT 60
QY 1655 ggcctcattccctatactcctaaactcctaataatgaagccttggcttctatgcatgtatgctcgt 1714
      1 GGTGCTCAATCTCAATATCTTAACCTTCATATGAAGACCTCTG-TTATGCATGATATGCTG 119
QY 1715 cttattatcccaatgactgctataactctctcttaagaagaactcagatgatataaag 1774
      120 CTTTATTCCCAATAGCTGCTATTAAGTCTCTCTTAAAGACACTGACATGTAATACAG 179
QY 1775 agtactgactgtctctctctcaataaattgcatlctatgactcagaagctgaatcaatc 1834

```

Db 180 AATATTCAGTCGTCCTTCCTCCAGATGATGTCATTTCATGATTCAGAGCTGATATTCATC 239

QY 1835 tcaagagagatggttttcattcccaagatcctctatgccatcccttggtttcttaccagttta 1894

Db 240 TCATATGATGATGGATTTCATCCAGATCTCTATGCGATACCTGATGCTTATGATATTTA 299

QY 1895 ctcattatattccactacacaaatatttccagcctctaggaalaccttactactt 1954

Db 300 CTCATGATATTCACATACCAAAATTTCTCCAGCTCTGGAGACCTTACATCTTGGAAT 359

QY 1955 cctcttccgcttctctgatttgg 1976

Db 360 CCTCTTCCATCTCTATTTGG 381

RESULT 11

LOCUS AM488854/c

DEFINITION AM488854 472 bp mRNA

VERSION UI-M-BH3-arz-a-05-0-01 s1 NIH-BMAP_M_S4 Mus musculus cDNA clone

KEYWORDS AM488854 GI:7059089

SOURCE EST.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 472)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene discovery

Genome Res. 6 (9), 793-806 (1996)

JOURNAL 97044477

MEDLINE

COMMENT Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel.: 301 443 1706

Fax: 301 443 9990

Email: mkt@mail.nih.gov

Oligo-dT track not found, Not 1 site shown in beginning of sequence

is likely internal to the message. cDNA library Preparation: M.B.

Soares lab Clone distribution: Researchers may obtain BMAP cDNA

clones from RESARCH GENETICS. It should be noted that Bento Soares

is generating a small number of additional specialized

non-redundant arrays of BMAP cDNAs whose availability will be

considered under appropriate and limited collaborative arrangements

Seq primer: M13 Forward

POLYA-No.

FEATURES

source

1..472

location/Qualifiers

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UI-M-BH3-arz-a-05-0-01"

/clone_11b="NIH-BMAP_M_S4"

/seq_stage="27-32 days"

/lab_host="PH10B (Life Technologies)"

/note="Vector: pTZ19-Pac (Pharmacia) with a modified

polylinker. Site 1: Not T; Site 2: Eco RI; The

NIH-BMAP_M_S4 library is a subtracted library of a series

ultimately derived from a mixture of individually tagged

normalized libraries from ten regions of the mouse brain

(cerebellum, brain stems, olfactory bulbs, hypothalamus,

cortex, amygdala, basal ganglia, pineal gland, striatum,

hippocampus) after a series of subtractions to reduce the

representation of cDNAs from which ESTs had already been

generated. The following serially subtracted libraries

were generated in this process: NIH-BMAP_M_S4,

NIH-BMAP_M_S3.3, NIH-BMAP_M_S3.2, NIH-BMAP_M_S3.1,

NIH-BMAP_M_S2, NIH-BMAP_M_S1. The subtracted library

(NIH-BMAP_M_S4) was constructed as follows: PCR amplified

Db 525 GAATATACCCAGAGAGTGTGATTTAGATGGGAGATTACAGATCAGGACAGAGGCT 584
Qy 2655 cctcagcattcccatctctctcaatatga 2682
Db 585 cctcagcattcccatctctctcaatatga 612

Search completed: June 26, 2001, 01:58:17
Job time: 2251 sec

GenInfo version 4.5
Copyright (c) 1994 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 08:00:53 ; Search time 24.32 Seconds

(without alignments)
896.877 Million cell updates/sec

Title: US-09-707-121-2

Perfect score: 4721

Sequence: 1 MEPLKDAHMAITFEASALP.....DQNKIKETGIIIPSPQI 893

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 241628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 241628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgnr2_6/prodata/2/1aa/5A.COMB.pep.*
2: /cgnr2_6/prodata/2/1aa/5A.COMB.pep.*
3: /cgnr2_6/prodata/2/1aa/6A.COMB.pep.*
4: /cgnr2_6/prodata/2/1aa/6B.COMB.pep.*
5: /cgnr2_6/prodata/2/1aa/PCUS.COMB.pep.*
6: /cgnr2_6/prodata/2/1aa/backlitest1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	268	5.7	633	4	US-08-557-006C-43
2	265	5.6	326	2	US-08-306-511A-10
3	265	5.6	326	2	US-08-893-274-10
4	265	5.6	326	3	US-08-581-418A-10
5	253	5.4	326	4	US-08-346-147B-10
6	253	5.4	326	4	US-08-822-936-10
7	253	5.4	326	5	PCT-US95-04636-10
8	252.5	5.3	257	2	US-07-857-224B-25
9	247.5	5.2	607	2	US-08-878-989-15
10	247.5	5.2	607	4	US-09-272-796-15
11	244	5.2	275	1	US-08-252-995D-13
12	244	5.2	275	2	US-08-834-108-13
13	242.5	5.1	445	3	US-09-101-146-1
14	235	5.0	372	2	US-08-837-593-9
15	231.5	4.9	647	3	US-09-031-563-5
16	231.5	4.9	648	3	US-09-031-563-5
17	231.5	4.9	729	2	US-08-677-298-2
18	231.5	4.9	1315	3	US-09-031-563-2
19	231.5	4.9	1315	3	US-09-031-563-25
20	231.5	4.9	1315	4	US-09-293-505-10
21	230.5	4.9	1001	4	US-09-060-410-2
22	228	4.8	257	4	US-09-101-146-6
23	228	4.8	1382	2	US-08-747-715-2
24	227	4.8	552	4	US-08-557-006C-40
25	226.5	4.8	618	2	US-08-770-761A-3
26	226.5	4.8	647	2	US-08-770-761A-8
27	226.5	4.8	660	2	US-08-770-761A-2

28	226.5	4.8	662	2	US-08-770-761A-5	Sequence 5, Appl
29	226.5	4.8	703	4	US-08-770-761A-7	Sequence 7, Appl
30	224.5	4.8	605	4	US-09-311-311C-26	Sequence 26, Appl
31	223.5	4.7	303	2	US-08-306-511A-9	Sequence 9, Appl
32	223.5	4.7	303	2	US-08-894-274-9	Sequence 9, Appl
33	223.5	4.7	603	4	US-09-196-122-2	Sequence 2, Appl
34	223.5	4.7	694	4	US-08-559-397A-31	Sequence 31, Appl
35	222.5	4.7	993	4	US-09-060-410-4	Sequence 4, Appl
36	222	4.7	278	2	US-07-857-224B-32	Sequence 32, Appl
37	221.5	4.7	316	1	US-08-403-634-31	Sequence 31, Appl
38	221.5	4.7	316	4	US-08-913-441B-31	Sequence 31, Appl
39	220.5	4.7	316	4	US-08-403-634-4	Sequence 4, Appl
40	220.5	4.7	316	4	US-08-913-441B-4	Sequence 4, Appl
41	220	4.7	506	1	US-08-369-780-2	Sequence 2, Appl
42	220	4.7	506	1	US-08-475-682-2	Sequence 2, Appl
43	220	4.7	506	1	US-08-780-833-2	Sequence 2, Appl
44	220	4.7	506	1	US-08-636-036-2	Sequence 2, Appl
45	220	4.7	506	3	US-08-918-509-2	Sequence 2, Appl

ALIGNMENTS

```

RESULT 1
US-08-557-006C-43
Sequence 43, Application US/08557006C
Patent No. 6258547
GENERAL INFORMATION:
APPLICANT: Berl, Rajindar K.
APPLICANT: Carling, David
APPLICANT: Ford, Robert A.
TITLE OR INVENTION: NUCLEIC ACID ENCODING AMP-ACTIVATED PROTEIN KINASE
FILE REFERENCE: NGAP/HRM37568/05T
CURRENT APPLICATION NUMBER: US/08/557,006C
CURRENT FILING DATE: 1996-03-06
PRIOR APPLICATION NUMBER: PCT/GB94/01093
PRIOR FILING DATE: 1994-05-20
PRIOR APPLICATION NUMBER: GB 9310489.1
PRIOR FILING DATE: 1993-05-21
PRIOR APPLICATION NUMBER: GB 9318010.7
PRIOR FILING DATE: 1993-08-11
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 43
LENGTH: 633
TYPE: PRT
ORGANISM: Yeast
FEATURE:
NAME/REV: PEPTIDE
LOCATION: (1)-(633)
OTHER INFORMATION: Yeast Snf1 polypeptide
US-08-557-006C-43

```

Query Match 5.7% Score 268; DB 4; Length 633;
Best local similarity 24.1% Pred. No. 5.1e-18;
Matches 92; Conservative 71; Mismatches 130; Indels 88; Gaps 13;

```

DB 107 LRLRLHPHILKLVDSKDE-1IMVIEVAGNELFDYIVQDKMSGEARFFOOLISAV 165
DB 105 QYMKRSHIVRAISPNHLLDRKGHIKAKGLVMTAGDGVFPFGYSYIAPVIAO 164
DB 166 EYGRHRIIVRDLPKPNLLDLHLVAKIADPGLSNTMDGFLKTSKGSNVAAPVIG 225
DB 165 GIEKTTTHMRSKPPISGSKRSYWSISGLITLDELVGRKLPUSLDPSERLKLFLITLDCVD 224
DB 226 KLY-----AGPEVWWSGVVILVYM-----LCRLPF-----DDF 255
DB 225 TLVILAE--ENGCLDIKELPEVIDLTKCLTFHPSKRPDPDELMKDKVFESEVSPLY-- 280
DB 256 SIPVLEKNISNGVYILPKFLISPGAGLILKRLMIVPDLNRLSIHRIWQIDMFVYDLPEYIL 315

```



```

PCT-US95-04636-10
: Sequence 10, Application PC/US9504636
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Cell-Cycle Regulatory Proteins, and Uses
: TITLE OF INVENTION: Related Thereto
: NUMBER OF SEQUENCES: 10
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII(Ext)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04636
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/246,147
: FILING DATE: 29-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/306,511
: FILING DATE: 14-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/248,812
: FILING DATE: 25-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/227,371
: FILING DATE: 14-APR-1994
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 326 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: N-terminal
: PCT-US95-04636-10

```

Query Match 5.4%; Score 253; DB 5; Length 326;

Best Local Similarity 28.7%; Pred. No. 5.3e-17;

Matches 93; Conservative 48; Mismatches 99; Indels 84; Gaps 16;

```

QY 25 GSNGLPTFNSIKILDRPOLKTIHPRLCOYVD---ISRKHK-RIYVAHECHESLSD 80
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 50 GERAMP--STIREVAVLHLETFEPHNVVRUPVCTSRDRETKLTLFEXVXODITT 107
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 81 LLREKRP---VSCVLAIAFVIOQLQYMNKGIYHRAISPHNILLDRKHILAKFGL 137
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 108 YL-DKVERGCVPTETIKDMFOLLRGLOELSHRYVHRLDKPQILITSSGQIKLADFL 166
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 138 YHHTAHDDVDPIFGPS-----YLAPRYAAGLIFKTHQMPKSKLPSPGSPDWSLIG 191
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 167 ART-----YSFQWALTIVVVTIMYRAPVILLQSSVAT-----PV-----DILMSVG 206
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 192 ILFELTCVGRKLFQ--SLDISERIKFLILLDCVDDTIVLAESGCLDI-----238
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 207 CFAEMERKRPDLRGSSVDVQGLK-----DIVGLGCEDEMPHVALPQAFHSKSA 259
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 239 -----IKELPETVILLKCLTFHPSKRPPTDELKMKVSESPPLTFTFPASIFSS 292
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 260 QPIEFVTDIDELGDKLLKCLTFNPKR-----ISAVSALSHRY-----PQD 302
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 293 SLRG---ADLTPEDISQKIDIN 313
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 303 LFRCKENLDSHLP--SQMTSEIN 324
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 8
US-07-857-2248-25
: Sequence 25, Application US/078572248
: Patent No. 5958784
: GENERAL INFORMATION:
: APPLICANT: Benner, Steven A.

```

: TITLE OF INVENTION: Predicting Folded Structures of Proteins
: NUMBER OF SEQUENCES: 114
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Steven A. Benner
: STREET: Hadlaubstrasse 151
: CITY: Zurich
: STATE: none
: COUNTRY: Switzerland
: ZIP: (note: this is an international post code) CH-8092
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
: COMPUTER: Apple Macintosh
: OPERATING SYSTEM: Macintosh 7.0
: SOFTWARE: Microsoft Word
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/857,224B
: FILING DATE: 03/25/92
: CLASSIFICATION: 436
: PRIOR APPLICATION DATA: none
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (international) 41 1 632 2830
: TELEFAX: (international) 41 1 262 2437
: TELEX: none
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 257
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE:
: DESCRIPTION: protein
: ORIGINAL SOURCE:
: ORGANISM: Saccharomyces cerevisiae
: FEATURE: Protein kinase; Table 8 Column 28
: PUBLICATION INFORMATION:
: AUTHORS:
: AUTHORS: Haaks, S. K.
: AUTHORS: Quin, A. M.
: AUTHORS: Hunter, J.
: TITLE: The protein kinase family
: JOURNAL: Science
: VOLUME: 241
: PAGES: 42-52
: DATE: 1988
: US-07-857-2248-25

```

Query Match 5.3%; Score 252.5; DB 2; Length 257;

Best Local Similarity 27.7%; Pred. No. 4.9e-17;

Matches 64; Conservative 50; Mismatches 84; Indels 33; Gaps 5;

```

QY 45 LKTIHPRLCOYVDISRKHRLIYVAHECHESLSDLLREKRPVSCVLAIAFVIOGL 104
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 55 LRLRHPIHILKIDYIKSKDE--IIMVIEYACNELPDYIVGRKMSQDEARPFQOIIISAV 113
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 105 QYMNKGIYHRAISPHNILLDRKHILAKFGLYHMTAHGDDVDPIGYPSTIAPEVIAQ 164
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 114 EYCHRIKRIYHRLKPNENLLDEHLNVKIAIDGSLNIMTQNLTKTSGSPNVAAPFVISG 173
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 165 GIKTTUHKPSKRPDLSPGSKVWSLGIILFELCVGRKLPQSLDISERIKFLILLDCVDD 224
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 174 KLV-----AGPEVIVWSCGVILYVM-----LKRRIAF-----DDE 203
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 225 TLVLAE--EHGCLDIKELPETVILLKCLTFHPSKRPPTDELKMKOVF 273
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 204 STIVLFKNISNGVITLPKFLSTGAAGLIKRLIYNPLNRISHIEMQDMPF 254
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 9
US-08-878-989-15
: Sequence 15, Application US/0887989
: Patent No. 5885803
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga

[illegible]

```

01b 414 GRRKSKNNGEGR--DEVSNTL VSA--LMTSTVCHJAFKFAVAFASCPVSTAFVAF 467
02 457 -----NLEFEGCSRGJGHRHSNLSLMDPTVLSYGLJRNKIKWGGAFVPILEDSNL 411
03b 468 INSEPTGLANSCKNPEEG-----LTVAVVFSALV ALBN1-----AFMFAACNVAVL 416
04 412 PRHSNNELSAATLPILEKPELV 447
05b 417 -----ADPELVVWSKVVIV 441
06
07 RESULT 10
08 OS-09-272-796-15
09 Sequence 15, Application US/09272796
10 Patent No. 6207148
11 GENERAL INFORMATION:
12 APPLICANT: Bandman, Olga
13 APPLICANT: Hillman, Jennifer L.
14 APPLICANT: Coffey, Noel G.
15 APPLICANT: Guogler, Karl G.
16 APPLICANT: Lal, Preeti
17 APPLICANT: Goll, Surya K.
18 APPLICANT: Shah, Parul
19 TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
20 TITLE OF INVENTION: KINASES
21 NUMBER OF SEQUENCES: 21
22 CORRESPONDENT ADDRESS:
23 ADDRESSEE: Lucyle Pharmaceuticals, Inc.
24 STREET: 4174 Forter Drive
25 CITY: Palo Alto
26 STATE: CA
27 COUNTRY: USA
28 ZIP: 94404
29 COMPUTER READABLE FORM:
30 MEDIUM TYPE: Diskette
31 COMPUTER: IBM Compatible
32 OPERATING SYSTEM: DOS
33 SOFTWARE: FastSeq for Windows Version 2.0
34 CURRENT APPLICATION DATA:
35 APPLICATION NUMBER: US/09/272-796
36 FILING DATE:
37 CLASSIFICATION:
38 PRIOR APPLICATION DATA:
39 APPLICATION NUMBER: 08/878,989
40 FILING DATE:
41 ATTORNEY/AGENT INFORMATION:
42 NAME: Billings, Lucy J J
43 REGISTRATION NUMBER: 36,749
44 REFERENCE/DOCKET NUMBER: PR-0121 US
45 TELECOMMUNICATION INFORMATION:
46 TELEPHONE: 415-855-0555
47 TELEFAX: 415-845-4166
48 TELEX:
49 INFORMATION FOR SEQ ID NO: 15:
50 SEQUENCE CHARACTERISTICS:
51 LENGTH: 607 amino acids
52 TYPE: amino acid
53 STRANDEDNESS: single
54 TOPOLOGY: linear
55 IMMEDIATE SOURCE:
56 LIBRARY: Genbank
57 CLONE: 1827450
58 OS-09-272-796-15
59
60 Query Match 5.28; Score 247.5; DB 4; Length 607;
61 Host Local Similarity 24.08; Prod. No. 9,9016;
62 Matches 107; Conservative 69; Mismatches 163; Indels 107; Gaps 18;
63
64 45 SIKLIGSGYLIKILPIIRLQGVVDISKG KHRGLVVAVER-----GRRSKL 79
65 50 AVXVTVDSKVAKPKQKELINLEFELNRDQDHRVPSRSHPEMDNINITYLELFSSKSLA 109

```

QY 80 DLLEKKPVSCSTVLCIAFVYUQIOTYNNKHGIVHVALSPHNLIDLRKHGILAKFGIYH 139
DB 110 HMKAKHILIEHYVYIKQILISIKYLIHQGLIHRLKGNFETIENNELVGVGCIATA 169
QY 140 MTAHGDVDPPI-GYPSYLAPEV-IAQGIKTDHMSKPKLPSCGSDVWSGLILFEL 197
DB 170 KLEPPQRKKTICGIPNVVAPEVLLRGS-----HGFPAVWSIGCVMTL 214
QY 198 CVAKKIFGSLDISEKRLTILDCVDTLLVLAERHGLDIKELPEYVIDLNNCLTFH 257
DB 215 LQSPPEFTADLKE-----TYKCI-----KOVHTFLASLSLPR-QLLAALIRAS 259
QY 258 PSKRPTPELIMKQVSESPLYTPTKPSALFSSSLKCADLILPEPISQLC----- 309
DB 260 PRPSPIDILKHDFETGCG--YTPDRLP--ISSCVTPDLTPPNPAKSLFAKVTKSLF 313
QY 310 --KDINNOYLAEHSIEHYVILMCLAGSDLEKILYNKELISKPTCTLP----- 356
DB 314 GRKKKKKNHADER--DEVSGI--VSG--LMTISVGHODARPEAPASCPAPVSLVETAF 367
QY 357 -----NLFEDGSEFGQGRDSSLLDPTVTLISQIYRNLKDVGGCAFYPLLEDQSNL 411
DB 368 DSSPQCTIASSGDFEG-----LIVATIVESALCALRNCI-----ATMPAFQGNPATL 416
QY 412 PHNSNNELSAATLPLIRKEDTFY 437
DB 417 -----AUPPEYVWSKWDY 431

RESULT 11

US-08-252-995D-13
Sequence 13, Application US/0825995D
Patent No. 5650501

GENERAL INFORMATION:

APPLICANT: Dennis, James W
APPLICANT: Heffernan, Mike
TITLE OF INVENTION: NOVEL SERINE/THREONINE KINASE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: HERESKIN & PARR
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/252-995D
FILING DATE: 02-JUN-1994
CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Kurdydk, Linda M
REGISTRATION NUMBER: 34,971
REFERENCE/DOCKET NUMBER: 3153-96
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 364-1398
INFORMATION FOR SPO ID NO: 13
SEQUENCE CHARACTERISTICS:
LENGTH: 275 amino acids
TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

ORIGINAL SOURCE:

ORGANISM: Saccharomyces cerevisiae
US-08-252-995D-13

Query Match 5.28; Score 244; DB 1; Length 275;
Best Local Similarity 26.88; Pred. No. 3-2e-16;
Matches 69; Conservative 51; Mismatches 95; Indels 42; Gaps 8;
QY 37 KILGFFQILKTIHTHRLDQVVDISKGNHFLVVAECER-SLEDDLHRRKPVSCSTVLC 95
DB 51 KLSLEIQLHKSMSHPNVQPTDFE-DNSNVYILFLPTGNSLMELLRRKRVTPPRF 109
QY 96 IAFEVYUQIOTYNNKHGIVHVALSPHNLIDLRKHGILAKFGIYHMTAGDVDPPI-GYP 154
DB 110 FTTOICGAIKYMHSRBYIHRLKIGNIFPDSNVNKAIGFGLAAVLAHSESKYTCGTP 169
QY 155 SYLAPEVIAQGIKTDHMSKPKLPSCGSDVWSGLILFELCVSKILFOSIDIS--E 211
DB 170 NYLAPEVL--MKHSHG-----SFEVDIWSIGVMYALLIGKPPVQAROVNTIYE 217
QY 212 RIKFILLTDCVDTLLVLAERHGLDIKELPEYVID-----LNNKTLFHRSKNYTPPEL 267
DB 218 RIK-----CRDSPRQKPIISDEKILIRQLISLDPTLEPSLLEI 257
QY 268 MKDQVSESPLYTPT 284
DB 258 MDYVWFKGTTPPSIPST 274

RESULT 12

US-08-834-108-13
Sequence 13, Application US/08834108
Patent No. 5976893

GENERAL INFORMATION:

APPLICANT: Dennis, James W
APPLICANT: Heffernan, Mike
TITLE OF INVENTION: NOVEL SERINE/THREONINE KINASE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: HERESKIN & PARR
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834-108
FILING DATE: 536

ATTORNEY/AGENT INFORMATION:

NAME: Kurdydk, Linda M
REGISTRATION NUMBER: 34,971
REFERENCE/DOCKET NUMBER: 3153-210
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 364-1398
INFORMATION FOR SPO ID NO: 13
SEQUENCE CHARACTERISTICS:
LENGTH: 275 amino acids
TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

ORIGINAL SOURCE:

ORGANISM: Saccharomyces cerevisiae
US-08-834-108-13

Query Match

5.28; Score 244; DB 2; Length 275;

Query Match 5.08; Score 235; DB 2; Length 372;

Best Local Similarity 24.5%; Pred. No. 4,56-15;

Matches 94; Conservative 62; Mismatches 145; Indels 82; Gaps 16;

```

QY 3 PLKAGKCAFTFFASALPHDVG-----SNGCLPTNSIKIKIGRQILKITHRLIQ 55
DB 34 PIRKIGKAVCIYGVSSVKNKINKEVALKKINNAFENRIDAIRTIIRKILRLHLENVLA 93
QY 56 YVDIISKGRHR---LVVAHECERSELEDLNKKPVSCSVLCIAEVLQSGYMNKIG 111
DB 54 LKDVMPHHRSEPKDYVLYVELMDIHLHOLIKSSQTLSDHQCYFQLJLGLKLYLSAN 153
QY 112 LVHALLSPHNLIDRRGHIKIAKHLYHMTAGUDVDPRICVPS---YLAPEVIAQIR 168
DB 154 LHHDLKPGNLLINAMCOKICDFGLAR-TSSGR-OPTMYVTVTRAFELLC----- 206
QY 169 TTDHPSKPELPSPKSDVWSLQIILFELVGRK-IPQSLDISREKFLIT-LDCVNDL 226
DB 207 CQDNY-----GTSDVWSVGCIFAEI-LGRKPFVPGTECLNQLKILINLQSGREED 257
QY 227 LVLAHEGCLDILKELPFT-----VILLNCLTFHRSKSPPTDELKXK 271
DB 258 LEFIDNKAKKRIKSLPYSGTPFSRLYHAPLALDLOMLVFDPSKRISYIEALQHP 317
QY 272 VFSEVSPLYPTPTKPSALPSSSLRCADLTLPEDISQLCKDINNLYLAERSLLEVYLMCL 331
DB 318 Y---MSPLYPNTDPPAOVIVLN-----DIDED-LGEETIRRM-KW-- 353
QY 354 AGGLEKEKLVKKELIRKPPIC 354
DB 354 -----SKLEYHPEAL 365

```

RESULT 15
US-09-011-563-7
Sequence 7, Application US/09011563A
Patent No. 6022708
GENERAL INFORMATION:
APPLICANT: Frederic de Sauvage
TITLE OR INVENTION: Fused
FILE REFERENCE: P1272
CURRENT APPLICATION NUMBER: US/09/011,563A
CURRENT FILING DATE: 1998-02-26
NUMBER OF SEQ ID NOS: 27
SEQ ID NO 7
LENGTH: 647
TYPE: PRT
ORGANISM: Homo sapiens
US-09-011-563-7

Query Match 4.9%; Score 231.5; DB 3; Length 647;

Best Local Similarity 22.0%; Pred. No. 2,86-14;

Matches 136; Conservative 82; Mismatches 202; Indels 197; Gaps 27;

```

QY 43 QILKTIHPRALCOYVDSRKGRLLVVAHEGERSLEDLREKRVSCSTYLCAFEVLQ 102
DB 53 EIMKGLHHPNIVHLD-SFTDKEVVAVTDVAEGHFLQILEDGKLPEDOVALAOLVS 111
QY 103 GLQYKNGKGIIVHALLSPHNLIDRRGHIKIAKHLYHMTAGUDVDPRICVPSYLAPEV 161
DB 112 ALVYLHSHRIITHRDMKPNILLAKGGIKLDFGPARAMSTNTWVLTSTIKGTPYMSPEL 171
QY 162 LAQILKFTTDHPSKPELPSPKSDVWSLQIILFELVGRKIFQSLDISREKFLITLDC 221
DB 172 VEE---KPYDH-----TADLMSVGCILYFRAVGTPPYATSTIFQLVSLIK--- 214
QY 222 VDDTLVLAEHSGCLDILKELPFTV---DLLNCLTFHRSKSPPT-PELEMKDKVSEV 276
DB 215 -----DPVR-WPSTISPCIKNPLQGLITKDPDRLMSWDLVYHPIAGIV 258

```

```

QY 277 S-----PLYPPTKQASLPSSSLRCADLTLPEDISQLCKD-----IN 313
DB 259 TITTEPAGFDISPTPTS-----RLPEEL-QVLKDDQAHRLAPKNGSRIL 304
QY 314 NDY--LAERSIEEYVYLMCLASGDEKELVKKELIRSKPPICLTPNLFEDGSHGQGRD 371
DB 304 QATKRAFEAMQKKHO---NTGPALEJEDKTSKVAPTAPLPRAL-----GATQ 349
QY 372 RSSLDDITVTLSLQILNRLKLDVGGFAVPHLLHQSNIPHSNNSNELSAATLPLIIR 411
DB 350 ESSSLAGI-----IASLTK-----SSWAKSGTGEVNSAR 379
QY 432 EKDTFQALNRIILFRLIKAVFYKKNOIKAEAVDIPFLMR-----GILWAALGVEGAI 486
DB 380 ENNTTPDCR-----AFPERPEVLQGRSTDVYDILNRPDSINKGHL----- 424
QY 487 HAKYDAIDKVTTPPTDROLEVDPICRHOVDLSSPEG-----HAKRHYLAAY-- 546
DB 425 -----ETTFVVP---IQLKAPL-----TILCNPPFCORIOSQLHFAQOILKGILEG 468
QY 537 VSH--PDIVYMQGIDSLCAPPLIYINFNNEALAYAGMSAFIPKYL----- 578
DB 469 ASHILDAFHYVLSLISSTSS-----DSVALYSPCRAGILGILLSLHRSQESNSLQD 521
QY 579 --YNFELKDNSHVIOEY 593
DB 522 SWYGFELDDMAVIOAY 538

```

Search completed: August 12, 2002, 09:04:24
Job time: 3811 sec

GenPac version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2001, 01:21:36 ; Search Time 3505.23 Seconds
(without alignments) 11835.030 Million cell updates/sec

Title: US-09-707-121-1

Perfect score: 2682
Sequence: 1 atgcttccctcgaagagacg.....tccatctctcgaatataga 2682

Scoring table: IDENTITY_MNC
Gapop 10.0 ; Gapext 1.0

Searched: 1344157 seqs, 773874588 residues

Total number of hits satisfying chosen parameters: 2688114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:
1: gb_ba1:
2: gb_ba2:
3: gb_ba3:
4: gb_ba4:
5: gb_ba5:
6: gb_ba6:
7: gb_ba7:
8: gb_ba8:
9: gb_ba9:
10: gb_ba10:
11: gb_ba11:
12: gb_ba12:
13: gb_ba13:
14: gb_ba14:
15: gb_ba15:
16: gb_ba16:
17: gb_ba17:
18: gb_ba18:
19: gb_ba19:
20: gb_ba20:
21: gb_ba21:
22: gb_ba22:
23: gb_ba23:
24: gb_ba24:
25: gb_ba25:
26: gb_ba26:
27: gb_ba27:
28: gb_ba28:
29: gb_ba29:
30: gb_ba30:
31: gb_ba31:
32: gb_ba32:
33: gb_ba33:
34: gb_ba34:
35: gb_ba35:
36: gb_ba36:
37: gb_ba37:
38: gb_ba38:
39: gb_ba39:
40: gb_ba40:
41: gb_ba41:
42: gb_ba42:
43: gb_ba43:

44: em_ov:
45: em_pa:
46: em_ph:
47: em_pi:
48: em_ro:
49: em_si:
50: em_sy:
51: em_un:
52: em_vl:
53: gb_sts1:
54: gb_sts2:
55: gb_sts3:
56: gb_sy:
57: gb_un:
58: gb_vl:
59: gb_vl2:
60: gb_vl3:
61: gb_vl4:
62: gb_vl5:
63: gb_vl6:
64: gb_vl7:
65: gb_vl8:
66: gb_vl9:
67: gb_vl10:
68: gb_vl11:
69: gb_vl12:
70: gb_vl13:
71: gb_vl14:
72: gb_vl15:
73: gb_vl16:
74: gb_vl17:
75: gb_vl18:
76: gb_vl19:
77: gb_vl20:
78: gb_vl21:
79: gb_vl22:
80: gb_vl23:
81: gb_vl24:
82: gb_vl25:
83: gb_vl26:
84: gb_vl27:
85: gb_vl28:
86: gb_vl29:
87: gb_vl30:
88: gb_vl31:
89: gb_vl32:
90: gb_vl33:
91: gb_vl34:
92: gb_vl35:
93: gb_vl36:
94: gb_vl37:
95: gb_vl38:
96: gb_vl39:
97: gb_vl40:
98: em_ba3:

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2659.6	99.2	2679	10	AX099148
2	1463.4	54.6	2428	89	AF161420
3	193	7.2	99878	77	AC087108
4	193	7.2	116822	83	AP002076
5	179.2	6.7	116530	83	AP001957
6	178.6	6.7	116530	83	AP001957
7	178.6	6.7	135584	66	AC020749
8	169.8	6.3	135584	66	AC020749

1	4	15	5.38	967.1	6.3	AT013873
2	10	156	5.4	41955.1	6.3	AT010447
3	11	145	5.4	17466	6.8	AT024695
4	12	147	5.1	72187	5.3	AT014343
5	13	113	4.2	101567	9.1	AT001820
6	14	91	4.3	52938	8.4	00081117
7	15	89	4.3	59843	8.4	000837815
8	16	80.5	3.8	38.5	9	AX032556
9	17	80.5	4.2	138759	6.3	AT014384
10	18	80.5	4.1	46463	6	CE034310
11	19	52.2	2.1	17418	10	154347
12	20	55.4	2.1	1359	12	AX03528
13	21	54.3	2.30	3413	9.6	MS05871
14	22	50.8	1.9	523	9.7	MS022456
15	23	50.5	1.9	1647	9.5	RM040819
16	24	50.5	1.9	1707	9.1	AB06K1
17	25	50.6	1.9	1863	8.5	AB022017
18	26	43.1	1.9	1808	6	00081500
19	27	43.4	1.8	224319	7.2	AT061717
20	28	43.4	1.8	2439	3	AT014473
21	29	43.2	1.8	1453	3	00080663
22	30	47.8	1.8	172475	8.8	AT025613
23	31	47.4	1.8	1076	5.3	NS080490
24	32	47.4	1.8	1341	10	AX084744
25	33	47.4	1.8	2219	12	AT044300
26	34	47	1.8	983	9.3	NS030335
27	35	47	1.8	2705	9.7	000806711
28	36	47	1.8	2754	9.7	000806711
29	37	46.5	1.7	1155	9.4	AT042482
30	38	46.5	1.7	2470	9.4	AT042483
31	39	46.4	1.7	1421	9	AX053632
32	40	46.4	1.7	2063	6	HY061018
33	41	46.4	1.7	2394	6	HY061018
34	42	46.4	1.7	70600	7.7	AT009078
35	43	46	1.7	1073	5.3	NS037775
36	44	46	1.7	4120	8	0010084
37	45	45.8	1.7	855	5.3	AX100846
38	46	45.8	1.7	855	5.3	AX100846
39	47	45.8	1.7	855	5.3	AX100846
40	48	45.8	1.7	855	5.3	AX100846
41	49	45.8	1.7	855	5.3	AX100846
42	50	45.8	1.7	855	5.3	AX100846
43	51	45.8	1.7	855	5.3	AX100846
44	52	45.8	1.7	855	5.3	AX100846
45	53	45.8	1.7	855	5.3	AX100846
46	54	45.8	1.7	855	5.3	AX100846
47	55	45.8	1.7	855	5.3	AX100846
48	56	45.8	1.7	855	5.3	AX100846
49	57	45.8	1.7	855	5.3	AX100846
50	58	45.8	1.7	855	5.3	AX100846
51	59	45.8	1.7	855	5.3	AX100846
52	60	45.8	1.7	855	5.3	AX100846
53	61	45.8	1.7	855	5.3	AX100846
54	62	45.8	1.7	855	5.3	AX100846
55	63	45.8	1.7	855	5.3	AX100846
56	64	45.8	1.7	855	5.3	AX100846
57	65	45.8	1.7	855	5.3	AX100846
58	66	45.8	1.7	855	5.3	AX100846
59	67	45.8	1.7	855	5.3	AX100846
60	68	45.8	1.7	855	5.3	AX100846
61	69	45.8	1.7	855	5.3	AX100846
62	70	45.8	1.7	855	5.3	AX100846
63	71	45.8	1.7	855	5.3	AX100846
64	72	45.8	1.7	855	5.3	AX100846
65	73	45.8	1.7	855	5.3	AX100846
66	74	45.8	1.7	855	5.3	AX100846
67	75	45.8	1.7	855	5.3	AX100846
68	76	45.8	1.7	855	5.3	AX100846
69	77	45.8	1.7	855	5.3	AX100846
70	78	45.8	1.7	855	5.3	AX100846
71	79	45.8	1.7	855	5.3	AX100846
72	80	45.8	1.7	855	5.3	AX100846
73	81	45.8	1.7	855	5.3	AX100846
74	82	45.8	1.7	855	5.3	AX100846
75	83	45.8	1.7	855	5.3	AX100846
76	84	45.8	1.7	855	5.3	AX100846
77	85	45.8	1.7	855	5.3	AX100846
78	86	45.8	1.7	855	5.3	AX100846
79	87	45.8	1.7	855	5.3	AX100846
80	88	45.8	1.7	855	5.3	AX100846
81	89	45.8	1.7	855	5.3	AX100846
82	90	45.8	1.7	855	5.3	AX100846
83	91	45.8	1.7	855	5.3	AX100846
84	92	45.8	1.7	855	5.3	AX100846
85	93	45.8	1.7	855	5.3	AX100846
86	94	45.8	1.7	855	5.3	AX100846
87	95	45.8	1.7	855	5.3	AX100846
88	96	45.8	1.7	855	5.3	AX100846
89	97	45.8	1.7	855	5.3	AX100846
90	98	45.8	1.7	855	5.3	AX100846
91	99	45.8	1.7	855	5.3	AX100846
92	100	45.8	1.7	855	5.3	AX100846
93	101	45.8	1.7	855	5.3	AX100846
94	102	45.8	1.7	855	5.3	AX100846
95	103	45.8	1.7	855	5.3	AX100846
96	104	45.8	1.7	855	5.3	AX100846
97	105	45.8	1.7	855	5.3	AX100846
98	106	45.8	1.7	855	5.3	AX100846
99	107	45.8	1.7	855	5.3	AX100846
100	108	45.8	1.7	855	5.3	AX100846
101	109	45.8	1.7	855	5.3	AX100846
102	110	45.8	1.7	855	5.3	AX100846
103	111	45.8	1.7	855	5.3	AX100846
104	112	45.8	1.7	855	5.3	AX100846
105	113	45.8	1.7	855	5.3	AX100846
106	114	45.8	1.7	855	5.3	AX100846
107	115	45.8	1.7	855	5.3	AX100846
108	116	45.8	1.7	855	5.3	AX100846
109	117	45.8	1.7	855	5.3	AX100846
110	118	45.8	1.7	855	5.3	AX100846
111	119	45.8	1.7	855	5.3	AX100846
112	120	45.8	1.7	855	5.3	AX100846
113	121	45.8	1.7	855	5.3	AX100846
114	122	45.8	1.7	855	5.3	AX100846
115	123	45.8	1.7	855	5.3	AX100846
116	124	45.8	1.7	855	5.3	AX100846
117	125	45.8	1.7	855	5.3	AX100846
118	126	45.8	1.7	855	5.3	AX100846
119	127	45.8	1.7	855	5.3	AX100846
120	128	45.8	1.7	855	5.3	AX100846
121	129	45.8	1.7	855	5.3	AX100846
122	130	45.8	1.7	855	5.3	AX100846
123	131	45.8	1.7	855	5.3	AX100846
124	132	45.8	1.7	855	5.3	AX100846
125	133	45.8	1.7	855	5.3	AX100846
126	134	45.8	1.7	855	5.3	AX100846
127	135	45.8	1.7	855	5.3	AX100846
128	136	45.8	1.7	855	5.3	AX100846
129	137	45.8	1.7	855	5.3	AX100846
130	138	45.8	1.7	855	5.3	AX100846
131	139	45.8	1.7	855	5.3	AX100846
132	140	45.8	1.7	855	5.3	AX100846
133	141	45.8	1.7	855	5.3	AX100846
134	142	45.8	1.7	855	5.3	AX100846
135	143	45.8	1.7	855	5.3	AX100846
136	144	45.8	1.7	855	5.3	AX100846
137	145	45.8	1.7	855	5.3	AX100846
138	146	45.8	1.7	855	5.3	AX100846
139	147	45.8	1.7	855	5.3	AX100846
140	148	45.8	1.7	855	5.3	AX100846
141	149	45.8	1.7	855	5.3	AX100846
142	150	45.8	1.7	855	5.3	AX100846
143	151	45.8	1.7	855	5.3	AX100846
144	152	45.8	1.7	855	5.3	AX100846
145	153	45.8	1.7	855	5.3	AX100846
146	154	45.8	1.7	855	5.3	AX100846
147	155	45.8	1.7	855	5.3	AX100846
148	156	45.8	1.7	855	5.3	AX100846
149	157	45.8	1.7	855	5.3	AX100846
150	158	45.8	1.7	855	5.3	AX100846
151	159	45.8	1.7	855	5.3	AX100846
152	160	45.8	1.7	855	5.3	AX100846
153	161	45.8	1.7	855	5.3	AX100846
154	162	45.8	1.7	855	5.3	AX100846
155	163	45.8	1.7	855	5.3	AX100846
156	164	45.8	1.7	855	5.3	AX100846
157	165	45.8	1.7	855	5.3	AX100846
158	166	45.8	1.7	855	5.3	AX100846
159	167	45.8	1.7	855	5.3	AX100846
160	168	45.8	1.7	855	5.3	AX100846
161	169	45.8	1.7	855	5.3	AX100846
162	170	45.8	1.7	855	5.3	AX100846
163	171	45.8	1.7	855	5.3	AX100846
164	172	45.8	1.7	855	5.3	AX100846
165	173	45.8	1.7	855	5.3	AX100846
166	174	45.8	1.7	855	5.3	AX100846
167	175	45.8	1.7	855	5.3	AX100846
168	176	45.8	1.7	855	5.3	AX100846
169	177	45.8	1.7	855	5.3	AX100846
170	178	45.8	1.7	855	5.3	AX100846
171	179	45.8	1.7	855	5.3	AX100846
172	180	45.8	1.7	855	5.3	AX100846
173	181	45.8	1.7	855	5.3	AX100846
174	182	45.8	1.7	855	5.3	AX100846
175	183	45.8	1.7	855	5.3	AX100846
176	184	45.8	1.7	855	5.3	AX100846
177	185	45.8	1.7	855	5.3	AX100846
178	186	45.8	1.7	855	5.3	AX100846
179	187	45.8	1.7	855	5.3	AX100846
180	188	45.8	1.7	855	5.3	AX100846
181	189	45.8	1.7	855	5.3	AX100846
182	190	45.8	1.7	855	5.3	AX100846
183	191	45.8	1.7	855	5.3	AX100846
184	192	45.8	1.7	855	5.3	AX100846
185	193	45.8	1.7	855	5.3	AX100846
186	194	45.8	1.7	855	5.3	AX100846
187	195	45.8	1.7	855	5.3	AX100846
188	196	45.8	1.7	855	5.3	AX100846
189	197	45.8	1.7	855	5.3	AX100846
190	198	45.8	1.7	855	5.3	AX100846
191	199	45.8	1.7	855	5.3	AX100846
192	200	45.8	1.7	855	5.3	AX100846
193	201	45.8	1.7	855	5.3	AX100846
194	202					

APPENDIX A1.1

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```


TITLE
JOURNAL
COMMENT

O'Connor, P., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Ketta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M.,
Roy, A., Santos, R., Schaefer, S., Schupbach, R., Seaman, S., Severy, P.,
Souque, C., Spencer, B., Stamp, Thompson, N., Stojanovic, N.,
Struss, N., Subramanian, A., Talamas, J., Testa, S., Theodorou, J.,
Travers, M., Travis, N., Trigglio, J., Vassiliou, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, C., Zainoun, J.,
Zemke, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (07-DEC-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All records were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WMR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L10976

Center clone name: 2421_M3

Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: dye terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960741

Consensus quality: 98992 bases at least Q40

Consensus quality: 99303 bases at least Q30

Consensus quality: 99520 bases at least Q20

Insert size: 100000; agarose-gel

Quality coverage: 12.5 in Q20 bases; agarose-gel

Quality coverage: 12.5 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 619; contig of 619 bp in length
* 720 719; gap of 100 bp
* 720 1345; contig of 626 bp in length
* 1346 1445; gap of 100 bp
* 1446 99878; contig of 98433 bp in length.

Location/Qualifiers

1. 99878

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone="C10 2421M3"

/clone_lib="C10 Human BAC"

1. 619

/note="assembly_fragment"

720. 1345

/note="assembly_fragment"

1446. 99878

/note="assembly_fragment"

BASE COUNT 31286 a 17710 c 17695 g 32986 t 201 others

ORIGIN

Query Match:

Best local similarity 7.2% Score 193; DB 77; Length 9878;

Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 atgttccctgaagaaacgaatgaagacgttaccctcttgctgaagctctgaacga 60
|||||
db 52475 ATGTTCCCTGAAGAACCACTCAATGGAGCGCTTACCTCTTCTGCCGCCGCGCCCA 52534

QY 181 aagqgaagacga 193
|||||
db 52655 AAGGGAAGCAAG 52667

RESULT 4

AP002076 116822 bp DNA HTG 26 AUG-2000
LOCUS
DEFINITION
Homo sapiens chromosome 4 clone 2230N19 map 4q22-q24. ***
SEQUENCING IN PROGRESS ***; 2 ordered pieces.

ACCESSION

AP002076 2 G1:9929910

VERSION

HTG; HTGS; PHASE2

KEYWORDS

Homo sapiens DNA, clone:2230N19.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;

1 (bases 1 to 116822)

REFERENCE

1 Tsai, S.F.

Submitted (19-MAY-2000) Shih-Peng Tsai, National Yang-Ming

University, Institute of Genetics, 155 Li-Kong St. Section 2,

Pelou, Taipei, Taiwan 11221, Republic of China

(E-mail: ympeisai@ym.edu.tw; URL: <http://genome.ym.edu.tw/>;

Tel:866-2-28267043; Fax:866-2-28264930)

On Aug 26, 2000 this sequence version replaced g1:7959557.

These sequences are draft human sequences, not finished sequences.

Those sequences are ordered pieces. And gaps between the contigs

of the same clone are represented as 100 N.

* NOTE: This is a 'working draft' sequence. It currently

consists of 2 contigs. Gaps between the contigs

are represented as runs of N. The order of the pieces

is believed to be correct as given, however the sizes

of the gaps between them are based on estimates that have

provided by the submitter.

* This sequence will be replaced

by the finished sequence as soon as it is available and

the accession number will be preserved.

* 1 38095; contig of 38095 bp in length

* 38096 38195; gap of 100 bp

* 38196 116822; contig of 78627 bp in length.

Location/Qualifiers

1. 116822

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4q22-q24"

/clone="2230N19"

BASE COUNT 37560 a 21375 c 20992 g 36795 t 100 others

ORIGIN

Query Match:

Best local similarity 7.2% Score 193; DB 83; Length 116822;

Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 atgttccctgaagaaacgaatgaagacgttaccctcttgctgaagctctgaacga 60
|||||
db 100663 ATGTTCCCTGAAGAACCACTCAATGGAGCGCTTACCTCTTCTGCCGCCGCGCCCA 100722

61 catgagatlttgaagaaacgaatgaagacgttaccctcttgctgaagctctgaacga 120
|||||
db 100723 CAGTGGATTTTGCGAAGCAAGAGACTCTCTCTGACACCAATTCATCAAAATTTTAAAG 100782


```

* 54406 54485: gap of unknown length
* 54486 55031: contig of 546 bp in length
* 55032 55111: gap of unknown length
* 55112 55534: contig of 423 bp in length
* 55535 55615: gap of unknown length
* 55615 56200: contig of 586 bp in length
* 56201 56280: gap of unknown length
* 56281 56635: contig of 355 bp in length
* 56636 56715: gap of unknown length
* 56716 57162: contig of 447 bp in length
* 57163 57242: gap of unknown length
* 57243 57906: contig of 664 bp in length
* 57907 57986: gap of unknown length
* 57987 58549: contig of 563 bp in length
* 58550 58629: gap of unknown length
* 58630 59379: contig of 750 bp in length
* 59459 59480: gap of unknown length
* 59480 60060: contig of 601 bp in length
* 60061 60140: gap of unknown length
* 60141 60378: contig of 238 bp in length
* 60379 60458: gap of unknown length
* 60459 60948: contig of 490 bp in length
* 60949 61028: gap of unknown length
* 61029 61701: contig of 673 bp in length
* 61702 61781: gap of unknown length
* 61782 62317: contig of 536 bp in length
* 62318 62397: gap of unknown length
* 62398 62948: contig of 551 bp in length
* 62949 63028: gap of unknown length
* 63029 63588: contig of 560 bp in length
* 63589 63669: gap of unknown length
* 63669 64225: contig of 557 bp in length
* 64226 64305: gap of unknown length
* 64306 64909: contig of 604 bp in length
* 64910 64989: gap of unknown length
* 64990 65664: contig of 675 bp in length
* 65665 65744: gap of unknown length
* 65745 66373: contig of 629 bp in length
* 66374 66453: gap of unknown length
* 66454 66965: contig of 512 bp in length
* 66966 67045: gap of unknown length
* 67046 67604: contig of 559 bp in length
* 67605 67684: gap of unknown length
* 67685 68233: contig of 549 bp in length
* 68234 68313: gap of unknown length
* 68314 68910: contig of 517 bp in length
* 68911 69419: gap of unknown length
* 69420 69499: gap of unknown length
* 69500 70097: contig of 598 bp in length
* 70098 70177: gap of unknown length
* 70178 70858: contig of 681 bp in length
* 70859 70938: gap of unknown length
* 70939 71544: contig of 606 bp in length
* 71545 71624: gap of unknown length
* 71625 72187: contig of 563 bp in length.

```

FEATURES

Query Match 5.18: Score 137; DB 63; Length 72187;
 Host Local Similarity 60.08; Pred. No. 176-27;
 Matches 249; Conservative 0; Mismatches 160; Indels 6; Gaps 1;

```

QY 1276 ccccccaatcagaaagagatcacagaaacaaatcaatattcctc 1335
| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 49733 CTGGCGTGGTATCGGCAAAAGACATCGATATCGTTGAGGCGGAGCTTTT 49792
| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1336 gacagctcttaagagctatccatataaadaaaccaatcgcgaagaagct 1395
| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 49793 GCGCGCTTCTGAGGGCTATCGGATACGCGGAGTACGTGAGCGGAGGAG 49852
| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1396 gacatcctcctcctcctcctcctcctcctcctcctcctcctcctc 1455
| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 49853 GATGTCGGCGGCTTACGAGGGGTCTTCATTTGGCGCGGCTCTCGAGCTTC 49906
| | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

QY 1456 attcatgccaaagtcagatgcaattgataaagacacccatcctcctacagatagaat 1515
| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 49907 CCCAATGGAGATTGCGGAGAGATGACAGAGTTTCGCTGAGAGAGATGAGATC 49966
| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1516 gaagtgatctcctcctcctcctcctcctcctcctcctcctcctcctc 1575
| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 49967 GAGGTGCAATATACCCCGCTGCGCATACAGTACAGTACAGTACAGTACAGTAC 50026
| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1576 gcaaaatttaagagtgatataaagcttgcgaagatgctcctcctcctcctcctc 1635
| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 50027 CCCAGCTCAGCCCGGCTGCTCCAGCCGCTGAGTCAACCGCCATGCTACCTACG 50086
| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1636 caagctctgactcactcttgctcctcctcctcctcctcctcctcctcctcctc 1690
| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 50087 CAGGACGAGATCTACTGAGGAGAGAGCTTCTCTATCTCAGCTTATACATGAG 50141
| | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 13

AP001820

AP001820 101947 bp DNA PRI 01-DEC-2000
 Homo sapiens genomic DNA, chromosome 4q22-q24, clone:2031A10.

DEFINITION complete sequence.

AP001820

AP001820.2 G1:11526580
 HTG.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Dec 1, 2000 this sequence version replaced q1:7594885.

PROJECT AV

Quality: the expected Phred/Phrap calculated error rate (per 10 kb)

is 0.0000; the estimated total number of errors is 0.0003.

FEATURES

source

1.101947

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4q22-q24"

/clone="2031A10"

HASH COUNT

34287 a 20004 c 19411 g 28245 t

ORIGIN

Query Match 4.28; Score 113; DB 91; Length 101947;

Host Local Similarity 100.0%; Pred. No. 186-17;

Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2570 aatttgaagctcactcttgcgaagatgaaatgaaatgaaatgaaatgaaatgaaatg 2629

DB 26799 AGTTTCACTCAGCTTGTGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 26858

QY 2630 ttaataaataaagctcactcctcctcctcctcctcctcctcctcctcctcctc 2682

DB 26859 TTAATAAATAAAGCTCAGCTCCTCAGCTCCTCAGCTCCTCAGCTCCTCAGCTCCTCAG 26911

RESULT 14

[illegible]

DB	16489	GACATCGACTGCTAGCGCAGCGGTTCCTATATCATCAATTACACTTAAACATGA	16438
RESULT:	15		
DMSB57K15			
LOCUS	DMSB57K15	59823 bp	DNA
DEFINITION	<i>Drosophila melanogaster</i> chromosome X clone BACD57K15 map 4B- ^a		HTG
KEYWORDS	strain y; cn bw sp; *** SHOUENICING IN PROGRESS ***; in unordered pieces.		
ACCESSION	A1121810	GI:6048213	
VERSION	HTG; HTGS_PHASE1.		
SOURCE	fruit fly.		
ORGANISM	<i>Drosophila melanogaster</i>		
REFERENCE	Fukayama, M.; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
AUTHORS	1 (bases 1 to 59823)		
TITLE	RA Murphy, L., Harris, D. and Baird, J.B.		
REMARK	Sequencing the X chromosome of <i>Drosophila melanogaster</i>		
REFERENCE	Unpublished		
AUTHORS	Sanger Centre		
TITLE	Wellcome Trust Genome Campus, Hinxton Hall, Hinxton,		
REMARK	Cambridge CB10 1SA, U.K.		
REFERENCE	2 (bases 1 to 59823)		
AUTHORS	Benois, P.		
TITLE	Direct Submission		
REMARK	Submitted (10-OCT-1999) European <i>Drosophila</i> Genome Sequencing Consortium		
COMMENT	This is a 'working draft' sequence. It currently consists of 119 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence.		
1	320: contig of 320 in length		
321	420: gap of unknown length		
421	559: contig of 239 in length		
559	759: gap of unknown length		
759	842: contig of 83 in length		
843	942: gap of unknown length		
943	1121: contig of 179 in length		
1122	1221: gap of unknown length		
1222	1250: contig of 29 in length		
1251	1350: gap of unknown length		
1351	1555: contig of 205 in length		
1556	1655: gap of unknown length		
1656	2002: contig of 347 in length		
2003	2102: gap of unknown length		
2103	2235: contig of 133 in length		
2236	2335: gap of unknown length		
2336	2591: contig of 356 in length		
2592	2791: gap of unknown length		
2792	3259: contig of 468 in length		
3260	3359: gap of unknown length		
3360	3542: contig of 283 in length		
3543	3742: gap of unknown length		
3743	4113: contig of 371 in length		
4114	4213: gap of unknown length		
4214	4601: contig of 388 in length		
4602	4701: gap of unknown length		
4702	4944: contig of 243 in length		
4945	5044: gap of unknown length		
5045	5482: contig of 438 in length		
5483	5582: gap of unknown length		
5583	5612: contig of 30 in length		
5613	5712: gap of unknown length		
5713	6202: contig of 490 in length		
6203	6302: gap of unknown length		
6303	6472: contig of 179 in length		
6473	6572: gap of unknown length		
6573	6752: contig of 180 in length		
6753	6852: gap of unknown length		
6853	7169: contig of 317 in length		

1710 1769: gap of unknown length
 1720 1769: count of 470 in length
 1740 1769: gap of unknown length
 1740 1769: count of 467 in length
 1807 1806: gap of unknown length
 1807 1806: count of 61 in length
 1868 1867: gap of unknown length
 1868 1867: count of 457 in length
 1868 1867: gap of unknown length
 1868 1867: count of 76 in length
 1901 1900: gap of unknown length
 1901 1900: count of 85 in length
 1901 1900: gap of unknown length
 1901 1900: count of 259 in length
 1925 1924: gap of unknown length
 1925 1924: count of 98 in length
 1925 1924: gap of unknown length
 1925 1924: count of 456 in length
 1939 1938: gap of unknown length
 1939 1938: count of 404 in length
 1944 1943: gap of unknown length
 1944 1943: count of 24 in length
 1966 1965: gap of unknown length
 1966 1965: count of 141 in length
 1966 1965: gap of unknown length
 1966 1965: count of 448 in length
 1967 1966: gap of unknown length
 1967 1966: count of 489 in length
 1967 1966: gap of unknown length
 1967 1966: count of 291 in length
 1967 1966: gap of unknown length
 1967 1966: count of 224 in length
 1967 1966: gap of unknown length
 1967 1966: count of 474 in length
 1967 1966: gap of unknown length
 1967 1966: count of 411 in length
 1967 1966: gap of unknown length
 1967 1966: count of 78 in length
 1967 1966: gap of unknown length
 1967 1966: count of 52 in length
 1967 1966: gap of unknown length
 1967 1966: count of 404 in length
 1967 1966: gap of unknown length
 1967 1966: count of 57 in length
 1967 1966: gap of unknown length
 1967 1966: count of 471 in length
 1967 1966: gap of unknown length
 1967 1966: count of 841 in length
 1967 1966: gap of unknown length
 1967 1966: count of 246 in length
 1967 1966: gap of unknown length
 1967 1966: count of 442 in length
 1967 1966: gap of unknown length
 1967 1966: count of 675 in length
 1967 1966: gap of unknown length
 1967 1966: count of 659 in length
 1967 1966: gap of unknown length
 1967 1966: count of 604 in length
 1967 1966: gap of unknown length
 1967 1966: count of 664 in length
 1967 1966: gap of unknown length
 1967 1966: count of 491 in length
 1967 1966: gap of unknown length
 1967 1966: count of 421 in length
 1967 1966: gap of unknown length
 1967 1966: count of 495 in length
 1967 1966: gap of unknown length
 1967 1966: count of 254 in length
 1967 1966: gap of unknown length
 1967 1966: count of 258 in length
 1967 1966: gap of unknown length
 1967 1966: count of 673 in length
 1967 1966: gap of unknown length

2269 2268: count of 274 in length
 2269 2268: gap of unknown length
 2307 2306: count of 469 in length
 2307 2306: gap of unknown length
 2307 2306: count of 817 in length
 2307 2306: gap of unknown length
 2307 2306: count of 614 in length
 2307 2306: gap of unknown length
 2307 2306: count of 456 in length
 2307 2306: gap of unknown length
 2307 2306: count of 154 in length
 2307 2306: gap of unknown length
 2307 2306: count of 298 in length
 2307 2306: gap of unknown length
 2307 2306: count of 90 in length
 2307 2306: gap of unknown length
 2307 2306: count of 549 in length
 2307 2306: gap of unknown length
 2307 2306: count of 426 in length
 2307 2306: gap of unknown length
 2307 2306: count of 114 in length
 2307 2306: gap of unknown length
 2307 2306: count of 497 in length
 2307 2306: gap of unknown length
 2307 2306: count of 254 in length
 2307 2306: gap of unknown length
 2307 2306: count of 695 in length
 2307 2306: gap of unknown length
 2307 2306: count of 141 in length
 2307 2306: gap of unknown length
 2307 2306: count of 447 in length
 2307 2306: gap of unknown length
 2307 2306: count of 431 in length
 2307 2306: gap of unknown length
 2307 2306: count of 447 in length
 2307 2306: gap of unknown length
 2307 2306: count of 949 in length
 2307 2306: gap of unknown length
 2307 2306: count of 81 in length
 2307 2306: gap of unknown length
 2307 2306: count of 742 in length
 2307 2306: gap of unknown length
 2307 2306: count of 74 in length
 2307 2306: gap of unknown length
 2307 2306: count of 833 in length
 2307 2306: gap of unknown length
 2307 2306: count of 517 in length
 2307 2306: gap of unknown length
 2307 2306: count of 1062 in length
 2307 2306: gap of unknown length
 2307 2306: count of 1172 in length
 2307 2306: gap of unknown length
 2307 2306: count of 1471 in length
 2307 2306: gap of unknown length
 2307 2306: count of 898 in length
 2307 2306: gap of unknown length
 2307 2306: count of 896 in length
 2307 2306: gap of unknown length

Query Match 4.48; Score 89; DB 84; Length 59821;

Host Local Similarity 58.58; Prev. No. 1.8e 11;

Matches 155; Conserved 0; Mismatches 110; Indels 0; Gaps 0;

1892 1892: count of 491 in length

1892 1892: gap of unknown length

1892 1892: count of 491 in length

1892 1892: gap of unknown length

1892 1892: count of 491 in length

1892 1892: gap of unknown length

•

•

•

GenScore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: June 26, 2001, 04:02:57 ; Search time 31.3 seconds

(without alignments)
977,321 Million cell updates/sec

Title: US-09-707-121-2

RefSeq score: 4721

Sequence: 1 MPKLMKDAEAGATFFASALP.....DGKINKIKPTGLTITPSDI 893

Scoring table:

BLAST/BLAST2
Gapop 10.0 ; Gapext 0.5

Searched: 93435 seqs, 3425466 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	286.5	6.1	611	1	SNF1_CANCA
2	282	6.0	619	1	SNF1_CANTR
3	280	5.9	1142	1	GIN4_YEAS
4	275	5.8	576	1	SNF1_SCHPO
5	274	5.8	620	1	SNF1_CANAL
6	268	5.7	634	1	SNF1_YEAST
7	266.5	5.6	326	1	CDK6_HUMAN
8	265.5	5.5	547	1	SPAK_HUMAN
9	252	5.3	512	1	K110_ARATH
10	252	5.3	705	1	CC5_YEAST
11	250	5.3	563	1	SPAK_RAT
12	249	5.3	683	1	PI3I_SCHPO
13	247	5.2	502	1	K111_SEGCE
14	246	5.2	548	1	AAK1_RAT
15	245	5.2	556	1	SPAK_MOUSE
16	249	5.1	443	1	MRK_ARATH
17	238.5	5.1	372	1	SPK1_SCHPO
18	237	5.0	371	1	MMK2_MOUSE
19	235.5	5.0	713	1	KP78_HUMAN
20	235	5.0	372	1	NTE1_TOBAC
21	231.5	4.9	1062	1	CC7_SCHPO
22	231	4.9	774	1	KEMK_MOUSE
23	231	4.9	1130	1	AAH1_HUMAN
24	229.5	4.9	603	1	PLK1_HUMAN
25	228.5	4.8	1518	1	KKX1_YEAST
26	228	4.8	1382	1	INSR_HUMAN
27	227.5	4.8	481	1	MSK_MOUSE
28	227	4.8	376	1	MPK4_ARATH
29	227	4.8	552	1	AAK2_HUMAN
30	227	4.8	552	1	AAK2_RAT
31	226.5	4.8	304	1	CDK4_HUMAN
32	226.5	4.8	915	1	KCQ4_YEAST
33	226	4.8	395	1	MPK6_ARATH

34	225.5	4.8	292	1	CC2H_ECTPI
35	225	4.8	296	1	CG2_DICDI
36	224.5	4.8	603	1	PLK1_MOUSE
37	224	4.7	376	1	MRK2_ARATH
38	223.5	4.7	384	1	MARK_PETRY
39	223.5	4.7	974	1	CG15_YEAST
40	222.5	4.7	714	1	HUNK_HUMAN
41	222.5	4.7	746	1	ABL_MLVAB
42	222.5	4.7	776	1	SN1L_RAT
43	222.5	4.7	1123	1	CDK4_MOUSE
44	221.5	4.7	360	1	CDK4_HUMAN
45	221.5	4.7	603	1	PLK1_RAT

ALIGNMENTS

RESULT	1	STANDARD	PRT	611 AA
SNF1_CANCA				
ID	000372			
AC	000372			
DI	01-NOV-1997 (Rel. 35, Created)			
DI	01-NOV-1997 (Rel. 35, Last sequence update)			
DE	15-JUL-1999 (Rel. 38, Last annotation update)			
DT	CARBON CATABOLITE DEPRESSING PROTEIN KINASE (EC 2.7.1 -)			
GN	SNF1			
OS	Candida glabrata (Yeast) (Torulopsis glabrata)			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida			
OX	NCBI_TaxID=5478;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NCIS84;			
KX	MEDLINE=97101049; PubMed=8945576;			
KA	Potter R., Kwon-Chung K.J.;			
RT	"Disruption of the SNF1 gene abolishes trehalose utilization in the			
RI	pathogenic yeast Candida glabrata."			
RL	Infect. Immun. 64:5269-5273(1996).			
CC	- FUNCTION: ESSENTIAL FOR RELEASE FROM GLUCOSE REPRESSION. IT			
CC	INTERACTS AND HAS FUNCTIONAL RELATIONSHIP TO THE REGULATORY			
CC	PROTEIN SNF4. COULD PHOSPHORYLATES CATH (BY SIMILARITY)			
CC	- SUBCELLULAR LOCATION: ASSOCIATED WITH THE NUCLEAR MEMBRANE (BY			
CC	SIMILARITY)			
CC	- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.			
CC	SNF1 SUBFAMILY.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL collaboration			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/databases/			
CC	or send an email to license@sib-sib.ch).			
CC				
DR	EMBL: L78130; AAB48642.1; -			
DR	ISSP: P24941; IAO1.			
DR	InterPro: IPR000719; -			
DR	InterPro: IPR002290; -			
DR	pfam: PF00069; PKINASE; 1.			
DR	PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.			
DR	PROSITE: PS00108; PROTEIN_KINASE_ST; 1.			
DR	PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.			
KW	transferase; Serine/threonine-protein kinase; ATP-binding;			
KW	phosphorylation; Carbohydrate metabolism; Nuclear protein.			
FT	DOMAIN	6	17	POLY-HIS.
FT	DOMAIN	39	290	PROTEIN KINASE
FT	NP_BIND	45	53	ATP (BY SIMILARITY)
FT	BINDING	68	68	ATP (BY SIMILARITY)
FT	ACT_SITE	161	161	BY SIMILARITY.
FT	MOD_RES	194	194	PHOSPHORYLATION (AUO-?) (BY SIMILARITY)
SQ	SEQUENCE	611 AA;	70049 MW;	89E17B12A4900C0D C9C64;

[illegible][illegible]

GN SNF1 OR GAT1 OR GCR1 OR PAS14 OR GIC2 OR YDR477W OR DB035.20.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycotales; Saccharomycotaceae; Saccharomycos.
 OX NCBI_TaxID=4932;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA MEDLINE-86289463; PubMed-3526554;
 RA Celenza J.L., Carlson M.;
 KI "A yeast gene that is essential for release from glucose repression
 encodes a protein kinase.";
 RL Science 233:1175-1180(1986).
 RN (2)
 RP SEQUENCE FROM N.A.
 RA Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E.,
 RA Horio A., Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
 RA Huntke-Smith S., Hyman R., Kemp C., Lashari D., Lew H., Lin D.,
 RA Mosedale D., Nakbata K., Namath A., Oelner P., Oh C., Patel F.X.,
 RA Roberts D., Schramm S., Schneider M., Shouren T., Shroff N.,
 RA Winant A., Yellon M., Botstein D., Davis R.W.;
 RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE OF 274-284; 528-539 AND 622-630, AND PHOSPHORYLATION SITE.
 RA MEDLINE-9431988; PubMed-7005477;
 RA Mitchellhill K.I., Stapleton D., Gao G., House C., Mitchell B.,
 RA Katsis F., Willets L.A., Kemp B.E.;
 KI "Mammalian AMP-activated protein kinase shares structural and
 functional homology with the catalytic domain of yeast Snf1 protein
 kinase.";
 RL J. Biol. Chem. 269:2361-2364(1994).
 CC -1- FUNCTION: ESSENTIAL FOR RELEASE FROM GLUCOSE REPRESSION. IT
 INTERACTS AND HAS FUNCTIONAL RELATIONSHIP TO THE REGULATORY
 PROTEIN SNF4. INTERACTS ALSO WITH S1P1, S1P2 AND GABR3. COULD
 PHOSPHORYLATES CASK.
 CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE NUCLEAR MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC SNF1 SUBFAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL contribution
 of the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M1371; AAA35058.1; -;
 DR EMBL: U33050; AAB64904.1; -;
 DR PIR: A26030; A26030.
 DR HSSP: P24941; LA01.
 DR SCD: S0002885; SNF1.
 DR InterPro: IPR000719; -;
 DR InterPro: IPR002290; -;
 DR Pfam: PF00069; PKinase; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS00108; PROTEIN KINASE ST; 1.
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.
 KW Phosphatase; Serine/threonine protein kinase; ATP-binding;
 KW Phosphorylation; Carbohydrate metabolism; Nuclear protein;
 FT DOMAIN 18 32 FOLD-HIS.
 FT DOMAIN 35 306 PROTEIN KINASE.
 FT NE_BIND 51 69 ATP (BY SIMILARITY).
 FT BINDING 84 84 ATP (BY SIMILARITY).
 FT ACT_SITE 177 177 HY SIMILARITY.
 FT MOD_RES 210 210 PHOSPHORYLATION (AUO-).
 SQ SEQUENCE 633 AA: 72045 MW: PSC63565C986C4E3 CRC64:

Query Match 5.7%; Score 268; OR 1; Length 633;
 Host Local Similarity 24.1%; Prod. No. 2.6e-10;
 Matches 92; Conservative 71; Mismatches 130; Indels 89; Gaps 13;
 45 LKTIHRLQYVDISRHERLVVAHCEPSLEDLREKRPVSCSTVLCIAFEVLGI 104

DB 107 LRLRLPHILIKYIVKIKKDE-TIVVTEYANLEFDYVGRKMSQGEARFFVITISAV 165
 QY 105 QVANKGIVHVALSHVNIILIRKCHIKIAKPCIYHMIAMQVDPDIPYSYLAPEVIAO 164
 DB 166 EYGHRIKIVHRLKPEINILLDEHILNVKIADEGSINIMTINKLEKSCSPYNAPEVISG 225
 QY 165 GIPEKTTIDMPKSKRPSPGSKDWSGIIILFELVVRKILFQSIDISERKTLITLWYDD 224
 DB 226 KIX-----AGPEYDVMSCVILVYM-----LCKRIIR-----IDR 255
 QY 225 TLIVLAIE-EMGSLDILIKELPEVIDLNGTLFHPSKRPDPDELAKKQVSEVSEPLY-- 280
 DB 256 SIPVLEKINSNVEYTLTPKFLSPGAAGLIKRMILVNPINRISIHIEIMQDWEFVDPEYLL 315
 QY 281 TPTPKASIFSSSLACAMILHPIHISQCMINNOYIAKSKIENYVYIMQIAQILKEKL 340
 DB 316 PEDLAFH-----PEENE-----NNDSKKGSSPD-----NDEIDNLL 348
 QY 341 VKKEIIRSKPPICTLPNLFEDGSEFGQDRSSLLDDTTVTLSCULRNRLKQVGEAF 400
 DB 349 VN-ILSS-----IMGYEKDEIVESILKSSHDIPAF-----NEIRDA----- 382
 QY 401 YLELDDSDNLPHSNNEIS 421
 DB 383 YMLIKENKSLIKDMKANKSVS 403

RESULT 7
 CDB6_HUMAN
 ID CDB6_HUMAN STANDARD; PRT: 326 AA.
 AC Q00534;
 DT 01-APR-1993 (REL 25, Created)
 DT 01-APR-1993 (REL 25, last sequence update)
 DT 01-OCT-2000 (REL 40, last annotation update)
 DE CELL DIVISION PROTEIN KINASE 6 (EC 2.7.1.-) (SERINE/THREONINE-PROTEIN
 KINASE PLSTIRE).
 GN CDB6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA MEDLINE-92347325; PubMed-1639063;
 RA Meyerson M., Anders G.H., Wu C.-L., Su L.-K., Gorka C., Nelson C.,
 RA Harlow E., Tsai L.-H.;
 RL "A family of human cdc2-related protein kinases.";
 RN EMBL J. 11:2909-2917(1992).
 RN (2)
 RP SOURCE OF 180-326 FROM N.A.
 RA Stanley A.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF COMPLEXES WITH INK4A & INK4D.
 RA MEDLINE-98421670; PubMed-9751050;
 RA Resse A.A., Tong L., Lee J.O., Jeffrey P.D., Pavletich N.P.;
 KI "Structural basis for inhibition of the cyclin-dependent kinase Cdk6
 by the tumor suppressor p16INK4a.";
 RL Nature 395:237-243(1998).
 CC -1- FUNCTION: PROBABLY INVOLVED IN THE CONTROL OF THE CELL CYCLE.
 CC INTERACTS WITH D-TYPE G1 CYCLINS.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC CDC2/CDBX SUBFAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL contribution
 of the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC

DB 329 TAELLKGFYK-----AKREYLEIK-----LTPPLAQBKKVRPSSGHL 376
 QY 418 AKRSLEHYIIMCAGDILKELVNKILISKPICTILNLEPHNG-STGGGRSSIL 376
 DB 377 --HKTEDMHWMS-----DDEMEKMS-----EKKAKFSDEKSRVKE 412
 QY 377 DDTITLSCGLERNLKWGSEAFPLEDDGNSLPHSSNNLSAATLPIIEKTE 436
 DB 413 ENPEIAVASSTIPPOLOS-----LSVHDSGPPNANHPYKASCAVNI-VIIRKSR 464
 QY 437 VOLNRI-----ILF-----DRLLKAVPPKKNOYKKAADVDPIMR 472
 DB 465 KELNDIREFTFGRDIDAGVSGELFSAVLVDKHDVIVANLQK-----IVDDPKLK 517
 QY 473 GLFWALLGVDA 485
 DB 518 LTFKIASGDOS 530

RESULT 9
 K110_ARATH STANFORD: PRT: 512 AA.
 AC Q38997: Q38976; 004728;
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE SNI-RELATED PROTEIN KINASE KIN10 (EC 2.7.1.-) (AKIN10).
 GN KIN10 OR SKIN10 OR 14P13.22.
 OS Arabidopsis thaliana (Mouse ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93010411; PubMed-1339373;
 RA Le Guen L., Thomas M., Blanchet M., Hallford N.G., Kreis M.;
 RT "Structure and expression of a gene from Arabidopsis thaliana
 encoding a protein related to SNI1 protein kinase.";
 RL Gene 120:249-254(1992).
 RN 12
 RP SEQUENCE FROM N.A.
 RX STRAIN-CV. COLUMBIA;
 RA Lessard P., Kreis M., Thomas M.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN 13
 RP SEQUENCE FROM N.A.
 RX STRAIN-CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Beaulieu M., Greasy T.H., Haas B., Wu D.,
 RA Bonding C.M., Koo H., Fujii C.Y., Dittelback T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN 14
 RP SEQUENCE OF 1-19 FROM N.A.
 RX STRAIN-CV. COLUMBIA;
 RX MEDLINE-95115691; PubMed-7816049;
 RA Le Guen L., Thomas M., Kreis M.;
 RT "Gene density and organization in a small region of the Arabidopsis
 thaliana genome.";
 RL Mol. Gen. Genet. 245:390-396(1994).
 RN 15
 RP FUNCTION: MAY PLAY AN IMPORTANT ROLE IN A SIGNAL TRANSDUCTION
 CC FUSION: REGULATE GENE EXPRESSION AND CARBOHYDRATE METABOLISM IN
 CC HIGHER PLANTS.
 CC 1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, SHOOTS AND LEAVES.
 CC 1- SIMILARITY: BELONGS TO THE SHK/THK FAMILY OF PROTEIN KINASES.
 CC SNI1 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@sib-sib.ch.
 CC
 DR EMBL: M93023; AAA32736.1;
 DR EMBL: X79707; CAA56146.1;
 DR EMBL: AC008261; AAF26165.1;
 DR EMBL: X94757; CAA64384.1;
 DR Mendel: 14016; ARATH:2326;tm14016.
 DR Mendel: 12898; ARATH:2098.1.
 DR InterPro: IPR000649;
 DR InterPro: IPR000719;
 DR InterPro: IPR001245;
 DR InterPro: IPR002290;
 DR Pfam: PF00627; UBA: 1.
 DR Pfam: PF00699; PKinase: 1.
 DR PRINTS: PR00109; TYRKINASE.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
 DR PROSITE: PS00108; PROTEIN_KINASE_SF_1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM_1.
 DR Transer: Serine/threonine-protein kinase; ATP-binding;
 KW Phosphorylation.
 KW DOMAIN 19 271 PROTEIN KINASE.
 FT NP_HIND 25 33 APP (BY SIMILARITY).
 FT HINDING 48 48 APP (BY SIMILARITY).
 FT ACT_SITE 142 142 BY SIMILARITY.
 FT MOD_RES 175 175 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 SQ SEQUENCE 512 AA: 58373 MW: 58186550AAASV6DF CNG64;
 Query Match 5.38; Score 252; Dh 1; length 512.
 Best Local Similarity 22.0%; Pred. No. 2.2e-09;
 Matches 85; Conservative 59; Mismatches 129; Indels 114; Gaps 7;
 QY 35 STKLIGR-----FOLLKITTHPLGCOYVVISGKRHRIVVAHCRSSLE 79
 DB 46 AKILINRRKIKNMDEHKVRREKILRLKPMHPTIKRYVITPTDITVMEVNSCHLP 105
 QY 80 DLREKRVGSCSTVLAJFAEVLQGLGYMKNKIVIRALSPNITLDRGRIKAFRIYH 139
 DB 106 DIVEKCHIQDFAKRNFGQILSGVEYCHNNVYHRIKPNILDSCKNKAIDPGISN 165
 QY 140 MAHGDVDVDFIGYPSVLAIPVYAGGFETIDHMSKRLSPGSCSDWSGIIITELCY 199
 DB 166 ITRDSHFLKTSQSPNTAFNEVLSKLY-----AGPEVDWSGCVLLVALLC 212
 QY 209 GRLKFOSLHSERKLFLLLEDCVDDTLVLAFFHGGTLIKRPEVITLANKDTPHPS 259
 DB 213 GRIPEPDENIPNLKKI-----KGCITYTSPSLSGARDLIPRMIVDPM 257
 QY 260 KRPTEDELMKRVSEVSPLYTPFKASLFSSSLCADLTLPEDISQGLDINNYLAE 319
 DB 258 KAVITPEIRQHMVQAHPLRYAP-----PPIVDQAKLIDETL-- 298
 QY 320 RSIERYVIMCLACDLKELVNKEILRSKPICTIPNLFEDGHSFGGGRKSLTDDT 379
 DB 299 -----QEVIN-----MGEDNILE-- 313
 QY 380 IVTLISGQIHNKIKDGVSEAFVPLED 406
 DB 414 -----SLRNTQNDXIVYVYIITDN 313

RESULT 10
 CC5_YEAST STANFORD: PRT: 705 AA.
 AC P32562;
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE CELL CYCLE PROTEIN KINASE CDC5/MSO2 (EC 2.7.1.-).
 GN CDC5 OR PKX2 OR MSO2 OR YMR001C OR YMR270.0YC.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.


```

00 BECOME EXCITING, EXPRESSED IN CHOROIO PLEXUS, DEVELOPING
01 MYOCARDIUM, PANCREATIC EPITHELIUM AND DORSAL ROOT GANGLIA.
02 -1- DOMAIN: PAPA BOX (PROLINE-ALANINE REPEATS) MAY TARGET THE KINASE
03 TO A SPECIFIC SUBCELLULAR LOCATION BY FACILITATING INTERACTION
04 WITH INTRACELLULAR PROTEINS SUCH AS ACTIN OR ACTIN-LIKE PROTEINS.
05 -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
06 STE20 SUBFAMILY.
07
08 This SWISS-PROT entry is copyright. It is produced through a collaboration
09 between the Swiss Institute of Bioinformatics and the EMBL outstation -
10 the European Bioinformatics Institute. There are no restrictions on its
11 use by non-profit institutions as long as its content is in no way
12 modified and this statement is not removed. Usage by and for commercial
13 entities requires a license agreement (see http://www.isb-sib.ch/announce/
14 or send an email to license@isb-sib.ch).
15
16 EMBL: 068190; AAC26000.1;
17 EMBL: AF068261; AAC24501.1;
18 EMBL: AF099990; AAC72239.1;
19 DR: HSSP: P00518; 1PKH.
20 InterPro: IPR000719;
21
22 Proam: P00069; PKINase; 1.
23 PROSITE: P500108; PROTEIN_KINASE_ST; FALSE_NGC.
24 PROSITE: P500107; PROTEIN_KINASE_ATP; 1.
25 PROSITE: P50011; PROTEIN_KINASE_DOM; 1.
26 Transferrase: Serine/threonine-protein kinase: ATP-binding.
27
28 DOMAIN 14 62 PRO/ALR-RICH.
29 FT DOMAIN 72 346 PROTEIN KINASE.
30 FT BINDING 78 86 ATP (BY SIMILARITY).
31 FT BINDING 101 101 ATP (BY SIMILARITY).
32 FT ACT_SITE 201 201 BY SIMILARITY.
33 FT DOMAIN 22 30 POLY-ALA.
34 FT DOMAIN 369 375 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
35 FT SITE 396 400 CASPASE CLEAVAGE RELATED SITE.
36 FT CONFLICT 11 11 V -> I (IN REF. 2).
37 FT CONFLICT 253 253 M -> I (IN REF. 2).
38 FT CONFLICT 403 403 E -> V (IN REF. 2).
39 SEQUENCE 553 AA: 60050 MW: 860158AEAF90C8 CRC64:
40
41 Query Match 5.3% Score 250; DB 1; Length 553;
42 Best Local Similarity 23.9%; Pred. No. 3; e-09;
43 Matches 118; Conservative 81; Mismatches 179; Indels 116; Gaps 20;
44
45 37 KILIPOLITIKIPRACQYVDISGKGRHRIVVAHCHENSLEMLR-----ERK--PV 88
46 114 ELKKIQANSQCSHPNVVYVYSPVKKDEHMLVKMLDSQSMIDILKYVNGCHKNVYL 173
47
48 89 SCSTVLCIAFEVLOGLOVKNKIGIVRALSPHNTLDRKGITKIAKGLYHMTAGHGDVD 148
49 174 EFAIMAILIKRVEIGLDITLHNGOJHRLKAGNILLGDSVOGLADPGVSAFIATGCVAT 233
50
51 149 FP-----IGYPSYLAPEVIAQGIKKTIDMPSKKPLSPGSKSWSTGILLFELCVGRK 202
52
53 234 RNVKRTFVGITPGWAPVEYMEQ--VRGYDF-----KADWSEFITALTELATGAA 280
54
55 203 LPQSLDISHKAFKFLITDQVDLITVLAEEHQCGLDKELPEVYDILNKLIFPHSKRP 262
56 281 PYKRYTPMVL--MLTLQMLPPLIFTVGRK---EMKKKKCKSKRLSLGLCKOKSKRP 335
57
58 263 TPDELKKQVSEVSPLYTPPTKPSLPSLSLKRADLTLPEDISGLCKDI-----NDYL 317
59 336 TAAELIKCKRFQK-----AKNREYVLEKLT-----LTPDPAQRAKKKRVYVSSCHL 383
60
61 318 AFPSIFEEVYIWLACQDLKELVNRKLIIRSKPPLITLPLNLPDGSFGQGRSLSD 377
62 384 --IKTDSQIMENS-----DDEMDKS-----EEGKA-AAASEKERRRKA 418
63
64 378 DTVVTLSTICQIKRRLKQVQ--EAYVYLLIEDQSNLHNSNNHLSAATLPIILREKOT 435
65 419 EENDEISV-----NAGGIIPOLOSLSVHDSQOPNANDYREGPCA--VNIIVLRHNS 469
66 436 EYOLNRI-----ILF-----DRLKAVPYKKNKIKREARVDIPILM 471

```

```

00 RESULT 12
01 PLOI SCHEP
02 ID PLOI SCHEP STANARI; PRT; 683 AA.
03 AC P50528;
04 DT 01-OCT-1996 (Rel. 34, Last sequence update)
05 DT 01-OCT-1996 (Rel. 34, Last sequence update)
06 DT 01-OCT-2000 (Rel. 40, Last annotation update)
07 DE SERINE/THREONINE-PROTEIN KINASE PLOI (EC 2.7.1.-).
08 GN PLOI OR SPAC23C11.16.
09 OS Schizosaccharomyces pombe (Eission yeast).
10 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
11 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
12 OC Schizosaccharomyces.
13 OX NCBI_TaxID=4896;
14 RN [1]
15 RP SEQUENCE FROM N.A.
16 RC STRAIN-972;
17 ME MEDLINE-95262899; PubMed-7744248;
18 RA Okura H., Hagan J.M., Glover D.M.;
19 RT "The conserved Schizosaccharomyces pombe Plo1, required to
20 form a bipolar spindle, the actin ring, and septum, can drive septum
21 formation in G1 and G2 cells."
22 genes Dev. 9:1059-1073(1995).
23 RN [2]
24 RP SEQUENCE FROM N.A.
25 RC STRAIN-972;
26 RA Brown D., Churcher C.M., Bartlett H.G., Kjaandream M.A., Wood V.;
27 RT Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
28 -1- FUNCTION: REQUIRED TO FORM A BIPOLAR SPINDLE, THE ACTIN RING AND
29 SEPTUM. FUNCTIONS UPSTREAM OF THE WHOLE SEPTUM FORMATION PATHWAY,
30 INCLUDING ACTIN RING FORMATION (REGULATED BY LATE SEPTATION GENES)
31 AND SEPTAL MATERIAL DEPOSITION (REGULATED BY EARLY SEPTATION
32 GENES). BEHAVES AS A "SEPTUM-PROMOTING FACTOR". AND COULD ALSO BE
33 INVOLVED IN INDUCING OTHER LATE EVENTS OF CELL DIVISION.
34 -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
35 CDC3/PLO SUBFAMILY.
36
37 This SWISS-PROT entry is copyright. It is produced through a collaboration
38 between the Swiss Institute of Bioinformatics and the EMBL outstation -
39 the European Bioinformatics Institute. There are no restrictions on its
40 use by non-profit institutions as long as its content is in no way
41 modified and this statement is not removed. Usage by and for commercial
42 entities requires a license agreement (see http://www.isb-sib.ch/announce/
43 or send an email to license@isb-sib.ch).
44
45 EMBL: X85758; GAA59766.1;
46 EMBL: Z98559; CAB11167.1;
47 HSSP: O63450; 1A06.
48 InterPro: IPR000719;
49 InterPro: IPR000959;
50 InterPro: IPR002290;
51
52 P1am: P500659; PLOI_Box; 2.
53 PROSITE: P500107; PROTEIN_KINASE_ATP; 1.
54 PROSITE: P500108; PROTEIN_KINASE_ST; 1.
55 PROSITE: P50011; PROTEIN_KINASE_DOM; 1.
56 Transferrase: Serine/threonine-protein kinase: ATP-binding.
57
58 DOMAIN 41 296 PROTEIN KINASE.
59 FT NP_BIND 47 55 ATP (BY SIMILARITY).
60 FT BINDING 69 69 ATP (BY SIMILARITY).
61 FT ACT_SITE 163 163 BY SIMILARITY.
62 FT DOMAIN 493 522 POLY-HOMOLOGY (PH2).
63 SEQUENCE 683 AA: 77301 MW: F11CH0EP9B913017 CRC64.

```


DEPLETED AND WHEN 5'-AMP RISES IN RESPONSE TO FUEL LIMITATION
AND/OR HYPOXIA. THIS IS A CATALYTIC SUBUNIT.
SUBUNIT: HEPHOMERIN OF A CATALYTIC SUBUNIT, A BETA AND A GAMMA
NON-CATALYTIC SUBUNIT.
-1- FIBROBLAST SPECIFICITY: LOW EXPRESSION IN KIDNEY, LIVER, LUNG, HEART,
AND BRAIN.
-1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
SNEI SUBFAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: 040819; AAC52355.1;
DR HSSP: 063450; 1A06;
DR InterPro: IPR000719;
DR InterPro: IPR002290;
DR Pfam: PF00069; pkinase.1;
DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1;
DR PROSITE: PS00108; PROTEIN_KINASE_ST.1;
DR PROSITE: PS00011; PROTEIN_KINASE_DOM.1;
KW Transferrase; Serine/threonine-protein kinase; Fatty acid biosynthesis;
KM Phosphorylation; ATP-binding; Multigene family;
KT DOMAIN: 16 268 PROTEIN KINASE;
FT NP_BIND: 22 30 ATP (POSSIBLE);
FT BINDING: 45 45 ATP (BY SIMILARITY);
FT ACT_SITE: 139 139 BY SIMILARITY;
FT MOD_RES: 172 172 PHOSPHORYLATION (AUTO-) (POTENTIAL);
SO SEQUENCE 548 AA; 62599 MW; 5CCA1281C195F867 CRC64;

Query Match 5.28; Score 246; DB 1; Length 548;
Best Local Similarity 25.18; Pred. No. 6e-09;
Matches 93; Conservative 55; Mismatches 125; Indels 98; Gaps 13;

Y 37 KILGRFQIKITTHRLACQVYVMSCKNHERLVVAH-----CRSLHDLIRK 86
D 60 KIRFONLKLFRPHHITKLVYSTSD-IFMYEVSGLGELDTCKNGLDEKESR 118
Y 87 PVSQSTVLCIAEVLQGLYANKKGYHRAISPINILDLKSHIKLAKGLYHRAIGDD 146
D 119 -----LFCQILSGVYCHRHVYHRODKPENVLLDAHNAKLAIDGILNMSSDGE 169
Y 147 VDEPIGVSYLAPDEVIAQIFKTDIMSKKPLSPGKSDVMSGLIIFELCVGRLEFS 206
D 170 LRTSGSPRYAPAEVLSKRY-----AGEVQIMSSGVILVALLOG----- 210
Y 207 LDISEMIKILILKCVDPHILVLAER-HQCLDIIRKIPRVIDLKNCLIFHPSKRP 264
D 211 -----LTFEDDHVPTLEKIKDGLTFYPOYLNPVSLSLKLMLQVDPMKRATI 259
Y 265 DELMKDQVSEVSPLQYFTPKPASFSSSLFCADLLPEDISQDKDINNDYLA----- 318
D 260 KQIRHFWKXQILPKY-----LPHVPSVSSIMIDPALKVCKPK 300
Y 319 ERSTEVYVTLKLAGDEKELVKKELRSPPGLTPNLFEDGSEFGGRD----- 371
D 301 ECSEEEV-LSC-----LYNR--NMDQPLAVVHLIDNRIMNEARDEVLTATSP 346
Y 372 KSLSLADIVT 382
D 347 RQSFLEDDHLLT 357

RESULT 15
SPAK_MOUSE STANDARD: PRT: 556 AA.
AT Q921W9;
DI 01-OCT-2000 (Rel. 40, Created)

01-OCT-2000 (Rel. 40, Last sequence update)
DI 01-OCT-2000 (Rel. 40, Last annotation update)
DE STE20/SPK1-RELATED PROLINE-ALANINE RICH PROTEIN KINASE (EC 2.7.1.-)
CN (STE-20 RELATED KINASE).
OS SPK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclerogastri; Muridae; Mus.
OX NCBI_TaxID:10090;
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE-Brain;
RX MEDLINE-20436889; PubMed-10980603;
RA Johnston A.M., Naselli G., Gomez L.J., Martin R.M., Harrison L.C.,
RA de Aizpura H.J.;
FT *SPAK, a STE20/SPK1-related kinase that activates the p38 pathway.*
FT Oncogene 19:4290-4297(2000)
CC -1- FUNCTION: MAY ACT AS A MEDIATOR OF STRESS ACTIVATED SIGNALS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (WHEN CASPASE-
CC CLEAVED) (PROBABLE).
CC -1- DOMAIN: PAPA BOX (PROLINE-ALANINE REPEATS) MAY TARGET THE KINASE
CC TO A SPECIFIC SUBCELLULAR LOCATION BY FACILITATING INTERACTION
CC WITH INTRACELLULAR PROTEINS SUCH AS ACTIN OR ACTIN-LIKE PROTEINS.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: AF099988; AAC72237.1;
DR HSSP: F00518; 1PRK;
DR MG0: MG1:1858416; Spak.
DR InterPro: IPR000719;
DR Pfam: PF00069; pkinase.1;
DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1;
DR PROSITE: PS00108; PROTEIN_KINASE_ST.1;
DR PROSITE: PS00011; PROTEIN_KINASE_DOM.1;
KW Transferrase; Serine/threonine-protein kinase; ATP-binding;
KT DOMAIN: 17 65 PRO/ALA-RICH;
FT DOMAIN: 75 349 PROTEIN KINASE;
FT NP_BIND: 81 89 ATP (BY SIMILARITY);
FT BINDING: 104 104 ATP (BY SIMILARITY);
FT ACT_SITE: 204 204 BY SIMILARITY;
FT DOMAIN: 22 31 POLY-ALA;
FT SITE: 372 378 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL);
FT SITE: 399 403 CASPASE CLEAVAGE RELATED SITE.
SO SEQUENCE 556 AA; 60319 MW; 66085A905431108 CRC64;

Query Match 5.28; Score 245; DB 1; Length 556;
Best Local Similarity 21.78; Pred. No. 7.1e-09;
Matches 117; Conservative 82; Mismatches 179; Indels 116; Gaps 20;

Y 37 KILGRFQIKITTHRLACQVYVMSCKNHERLVVAHCRSLHDLIRK ----ERK-IV 88
D 117 ELKELQAMSGSSEHNVVYTFYFVVDKELMVKLLSGSSMLDIKLYIVNGEKHNVL 176
Y 89 SCSTVLCIAEVLQGLYANKKGYHRAISPINILDLKSHIKLAKGLYHRAIGDD 146
D 177 HRAIILATILKEVILQGLYHNGOIHNDIKACNILLGRASVOIALPGSAPALAGIVT 236
Y 149 EP-----IGPSTLAPRVIAQIFKTDIMSKKPLSPGKSDVMSGLIIFELCVGR 202
D 237 RKKVKKTFVGTCPKAPVPMQ--VRGYD-----KADMSQCIYALIELADAA 283
Y 203 LPQSLDIHRIKFLATILKCVDPHILVLAERHGLDIIRKIPRVIDLKNCLIFHPSKRP 262
D 284 PYHXYPPMKVL--MQLTONDPPILETGVDEK-----EMKKYKGSFKLLSLGDKPSKRP 338

KM Ischemic disorder; inflammation; diabetes mellitus; fibrosis; mitosis;
 KM mesangial disorder; growth regulation; wound healing; T cell activation;
 KM immunosuppressant.
 OS Homo sapiens.
 W09953036-A2.
 PD 21-OCT-1999.
 PD 13-APR-1999; 99WO-0808150.
 PE 14-APR-1998; 98US-0081784.
 PA (SUGEN) SUGEN INC.
 PI Plowman G, Martinez R, Whyte D;
 PS WPI: 1999-611301/52.
 DK N-PSDB: AAZ40480.
 PT Novel kinase-related polypeptides used for the diagnosis and treatment
 of kinase-related diseases and disorders.
 PS Claim 11: Page 260-262; 387pp; English.

XX This sequence represents a novel STE20-related protein kinase. The
 CC invention relates to nucleic acid molecule encoding a kinase polypeptide
 CC selected from STK2K, STK3, STK4, STK5, STK6, STK7, ZC1, ZC2, ZC3,
 CC ZC4, KHS2, S0U01, S0U03, CK2, PAK4 and PAK5. The proteins are used to
 CC identify agonists and antagonists, and to raise antibodies. The
 CC polynucleotides are useful in gene therapy protocols. The polynucleotides,
 CC polypeptides, antibodies, antagonists and agonists may be used to treat
 CC diseases such as immune-related disorders and diseases (e.g. rheumatoid
 CC arthritis, atherosclerosis, chronic inflammatory bowel disease (e.g.
 CC Crohn's disease), asthma, osteoarthritis, psoriasis, atherosclerosis,
 CC thinitis, autoimmunity, and organ transplantation, chronic inflammatory
 CC disease, multiple sclerosis, organ transplantation, myocardial
 CC infarction, cardiovascular disease, stroke, renal failure, oxidative
 CC stress-related neurodegenerative disorders (e.g. amyotrophic lateral
 CC sclerosis, Parkinson's disease and Leigh syndrome), cancer,
 CC cardiomyopathies, ischemic disorders, inflammatory disorders, diabetes
 CC mellitus, fibrotic and mesangial disorders. The proteins may also be
 CC useful for cell growth regulation (e.g. in wound healing), T cell
 CC activation, mitosis control, and as immunosuppressants.

XX Sequence 516 AA:

Query Match 5.5% Score 260.5; DH 20; Length 516;

Best Local Similarity 23.3%; Pred. No. 11e-15; Mismatches 113; Gaps 18;

Matches 115; Conservative 86; Indels 113; Gaps 18;

37 KILGRQLTKTHPEICVYDLSGKHEKVVVAHFCENSLDLK-----ERK-PV 88
 76 CLKIGAMSGSHPVYFYSYVKDELIVMKLISGSMIDIKYIVNGKHNYL 135
 89 SCSTVLCIAFEVYLGLOVYNNKHGIVYFALSPHNLIRKRIKAKGKILHMAHDDDD 148
 136 CQALIACTIKKEVQDYHINQYHRAIKAGNLIQDSVQVADYGVSAFTARGDPT 195
 149 FP-----IGYTSYLAPEVYIAQIKRIIDHMSKKRPIPSGKSHVSLGILIPKICVGRK 202
 196 ENKVFILVQCPQWAPVMEQ--VRYGDI-----KADWELQALALAQDA 242
 203 LFGSLDISERLKFLTIDCVDDILVLAERHGLDILKEIPETVLDLANKCLTFHPSKRP 262
 243 PYYKKPMKVL--MLTQIDPFTCTGYECK--CMKKKYSKSTKILSLICIDKQPSKP 297
 263 TPDLEMKDKVESEVSPLYPTKPSALFSSSLKCAVDLTLPEDISQLCKDI-----NNDYL 317
 298 TAACTIKKCTFGK-----AKREYIIKEL-----LTPDIDAGRAKKVRYGSSGSHL 345

QY 318 AKRSIEHYVYIWCIAQDILKELVNKEITNSKPPICITLNPFLPQGE-SFGSGRKRSSL 376
 DB 346 --hkledqaws-----dlemqks-----eeqkaalsekstrvke 381
 QY 377 DDTIVTLSCQLRNLKDVGEAFPLEDDQSNLPHSNNELSAMATPLIREKDE 436
 DB 382 CQPIAVSASTYPEQIG-----LSVHDSQGPANPDYFCASSCAVNIIVIRIS 433
 QY 437 YQENRI-----ILF-----DRLKAVPYKKNOIMKEANVDIPLMR 472
 DB 434 KQINDIRFEIPQDAVSGELISQVDAQHVVLAANLYK-----LVDDPKAIK 486
 QY 473 GTFMAALGVNCA 485
 DB 487 LLIKIASQDGS 499

RESULT 7

AAV01496 standard; peptide: 520 AA.

AAV01496:

21-MAY-1999 (first entry)

Human pan-s/tk-1A receptor polypeptide.

Pancreatic serine/threonine kinase receptor: pan-s/tk: cell growth;

differentiation; gene therapy; ligand: screening: cell proliferation;

regulator: pancreatic.

Homo sapiens.

Key Location/Qualifiers

FT Misc-difference 11 /note="unspecified"

W09907854-A2.

18-FEB-1999.

11-AUG-1998; 98WO-US16640.

11-AUG-1997; 97US-0909354.

(ONTO-) ONTOGENY INC.

Barker DD, Miao N, Pang K;

WPI: 1999-167430/14.

N-PSDB: AAZ6148.

New pancreatic serine/threonine kinase (pan-s/tk) receptor

polypeptides - useful for modulating cell proliferation,

differentiation and survival in animals

Claim 1: Page 66-67; 77pp; English.

The invention relates to isolated and/or recombinant pancreatic serine/
 CC threonine kinase receptor (pan-s/tk) polypeptides. Sequences of rat and
 CC human pan-s/tk polypeptides and nucleic acids encoding them are provided.
 CC Host cells containing vectors comprising the pan-s/tk genes are
 CC used for the recombinant expression of the proteins. The pan-s/tk
 CC polypeptides are useful for modulating cell growth, differentiation or
 CC survival in an animal cell. The polynucleotides are also useful for
 CC detecting pan-s/tk ligand on cells in samples and for gene therapy.
 CC ligands identified, together with soluble polypeptides, are useful for
 CC screening test compounds that modulate the bioactivity of a pan-s/tk
 CC receptor. The discovery of this new class of pan-s/tk polypeptides
 CC provides an alternative method of mediating cell proliferation,
 CC differentiation and survival in animals. The polypeptides also represent
 CC the first identified specific regulators of pancreatic development. The
 CC present sequence represents a human pan-s/tk-1A receptor polypeptide.

XX Sequence: 520 AA:
 Query Match: 5.5%, Score: 260.5, DB: 20, Length: 520;
 Best Local Similarity: 24.3%, Pred. No. 1, 1e-15;
 Matches: 115; Conserved: 86; Mismatches: 179; Indels: 11; Gaps: 10;

UY 47 KILGKPVLLKTLHPRLCYVYDLSRGRRLVAVARERESLDELRLK -----ERK PV 88
 DB 80 GLLKQVPMSPSHQVAVYVYTVYVKKGLWLVWKLISQSMGLDKLYLVYVQKQVYL 139
 UY 89 SCSTVLCIAFEVLCGLVYNNKRGIVHVALSPNILLRKGHLKLAEGLYHMTARDVD 148
 DB 140 CCAATLALIKVYVGLVYVHNGHTRHDKAGHLLIQGASVQVADTYSATLATQVDT 199
 UY 149 EP -----IGYSVYLAPEVLAAGLIPKTDHMSKKRLDPSQSDWSLALIFELVGRK 202
 DB 200 RUKVKTLVQVPMAPVPMVQ VYVYDT -----Kadmsqglatalaqa 246
 UY 203 LFGSLDLSERKELLFLKVPDLVLAELHGTDLIKELPEVLDLNLKTLFPHSKRP 262
 DB 247 PYHKYPMKVLMLLIPDPTLCTVQVK -----Cemkkyqsksttkllslclqkpskrp 301
 UY 263 TFEELMKQKVESEVSPLYPTPTKPSLFSSSLRCALDLPEDLSQVLCVLI -----NNYL 417
 DB 402 LKGLIKKKTGK -----akrsgylckl -----lttptdgaqakvrrvpssschl 449
 UY 418 AKRSLEFVYVYVWLAAGLIPKTLVNSKRPGLTLPNPLFGRPE SPQSDQSSSL 376
 DB 450 LKGLHAWES -----ademeks -----ecckaklsqsksttkvke 385
 UY 457 DDFVYVLSLQVNNKLVGGLVAVYVPLLEDDQSNLPSSNSNNELSAATLPLIKKOTE 436
 DB 486 CQPLQVAVASSTIPQVIGS -----lsvdsqtppmadcytsssgvnlvlllntst 437
 UY 487 YOLNRI -----LLF -----DLIKAVYKRNQVKEAVDTPELMR 472
 DB 488 Kellndtfttprdtadgvsqclfsaqlvdgldvltvaanlqk -----lvddpklk 490
 UY 491 TLTKLASQDS 503

RESULT 8
 ID AAY01497 standard; Protein: 520 AA.
 AC AAY01497:
 AA MAY 1999 (first entry)
 GI 21 Human pan s/tk 1b receptor polypeptide.
 FE Human pan s/tk 1b receptor polypeptide.
 KM Differentiation gene therapy; ligand: secretin; cell proliferation;
 KM regulation: pancreatic.
 XX Human secretin.
 OS
 EN W090705A A2.
 ID 18 FEB 1999.
 XX
 FE 11 AUG 1998; 9806 0516640
 DB 11 AUG 1997, 9705 0909 614.
 XX
 PA (Genetic) ONTODENY INC.
 XX
 FI Backer ID, Mayo N, Band K.
 XX

DR WPI: 1999-167430/14
 DR N PUB: AMX26149.
 XX
 PT New pancreatic serine/threonine kinase (pan-s/tk) receptor
 PT polypeptides - useful for modulating cell proliferation.
 PT differentiation and survival in animals
 XX
 PS Claim 1: Page 70-71: 77pp; English.
 CC The invention relates to isolated and/or recombinant pancreatic serine/
 CC threonine kinase receptor (pan-s/tk) polypeptides. Sequences of rat and
 CC human pan-s/tk polypeptides and nucleic acids encoding them are provided.
 CC Host cells containing vectors comprising the pan-s/tk genes are
 CC used for the recombinant expression of the proteins. The pan-s/tk
 CC polypeptides are useful for modulation of cell growth, differentiation or
 CC survival in an animal cell. The polynucleotides are also useful for
 CC detecting pan-s/tk ligand on cells in samples and for gene therapy.
 CC Ligands identified, together with soluble polypeptides, are useful for
 CC screening test compounds that modulate the bioactivity of a pan-s/tk
 CC receptor. The discovery of this new class of pan-s/tk polypeptides
 CC provides an alternative method of modulating cell proliferation,
 CC differentiation and survival in animals. The polypeptides also represent
 CC the first identified specific regulators of pancreatic development. The
 CC present sequence represents a human pan-s/tk 1b receptor polypeptide.

Sequence: 520 AA:
 Query Match: 5.5%, Score: 260.5, DB: 20, Length: 520;
 Best Local Similarity: 24.3%, Pred. No. 1, 1e-15;
 Matches: 115; Conserved: 86; Mismatches: 179; Indels: 11; Gaps: 10;

UY 47 KILGKPVLLKTLHPRLCYVYDLSRGRRLVAVARERESLDELRLK -----ERK PV 88
 DB 80 GLLKQVPMSPSHQVAVYVYTVYVKKGLWLVWKLISQSMGLDKLYLVYVQKQVYL 139
 UY 89 SCSTVLCIAFEVLCGLVYNNKRGIVHVALSPNILLRKGHLKLAEGLYHMTARDVD 148
 DB 140 CCAATLALIKVYVGLVYVHNGHTRHDKAGHLLIQGASVQVADTYSATLATQVDT 199
 UY 149 EP -----IGYSVYLAPEVLAAGLIPKTDHMSKKRLDPSQSDWSLALIFELVGRK 202
 DB 200 RUKVKTLVQVPMAPVPMVQ VYVYDT -----Kadmsqglatalaqa 246
 UY 203 LFGSLDLSERKELLFLKVPDLVLAELHGTDLIKELPEVLDLNLKTLFPHSKRP 262
 DB 247 PYHKYPMKVLMLLIPDPTLCTVQVK -----Cemkkyqsksttkllslclqkpskrp 301
 UY 263 TFEELMKQKVESEVSPLYPTPTKPSLFSSSLRCALDLPEDLSQVLCVLI -----NNYL 417
 DB 402 LKGLIKKKTGK -----akrsgylckl -----lttptdgaqakvrrvpssschl 449
 UY 418 AKRSLEFVYVYVWLAAGLIPKTLVNSKRPGLTLPNPLFGRPE SPQSDQSSSL 376
 DB 450 LKGLHAWES -----ademeks -----ecckaklsqsksttkvke 385
 UY 457 DDFVYVLSLQVNNKLVGGLVAVYVPLLEDDQSNLPSSNSNNELSAATLPLIKKOTE 436
 DB 486 CQPLQVAVASSTIPQVIGS -----lsvdsqtppmadcytsssgvnlvlllntst 437
 UY 487 YOLNRI -----LLF -----DLIKAVYKRNQVKEAVDTPELMR 472
 DB 488 Kellndtfttprdtadgvsqclfsaqlvdgldvltvaanlqk -----lvddpklk 490
 UY 491 TLTKLASQDS 503

RESULT 9
 ID AAY01517 standard; Protein: 512 AA.
 XX

AC: AAG36157;
XX
XX 18-JUN-2000 (first entry)
DE: Arabidopsis thaliana protein fragment SEQ ID NO: 44272.
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
XX hybridisation assay: genetic mapping; gene expression control; promoter;
XX termination sequence.
CS Arabidopsis thaliana.
XX
XX EPI034405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000HP-0301439.
XX
XX 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-0121180.
PR 09-MAR-1999; 9905-0123548.
PR 24-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 10-APR-1999; 9905-0128845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130447.
PR 23-APR-1999; 9905-0130510.
PR 28-APR-1999; 9905-0130891.
PR 30-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 04-MAY-1999; 9905-0132407.
PR 05-MAY-1999; 9905-0132484.
PR 06-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 07-MAY-1999; 9905-0132487.
PR 11-MAY-1999; 9905-0132863.
PR 14-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 18-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.

PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 02-JUL-1999; 9905-0142154.
PR 06-JUL-1999; 9905-0142055.
PR 08-JUL-1999; 9905-0142390.
PR 09-JUL-1999; 9905-0142803.
PR 12-JUL-1999; 9905-0142920.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145089.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 03-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 05-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149375.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.

```

26 AUG-1999; 9905-0150884
PR 27 AUG-1999; 9905-0151065
PR 27 AUG-1999; 9905-0151066
PR 27 AUG-1999; 9905-0151066
PR 30 AUG-1999; 9905-0151304
PR 31 AUG-1999; 9905-0151408
PR 01 SEP-1999; 9905-0151900
PR 07 SEP-1999; 9905-0152464
PR 10 SEP-1999; 9905-0154070
PR 13 SEP-1999; 9905-0154758
PR 15 SEP-1999; 9905-0154018
PR 20 SEP-1999; 9905-0154039
PR 22 SEP-1999; 9905-0154779
PR 23 SEP-1999; 9905-0155159
PR 24 SEP-1999; 9905-0155486
PR 24 SEP-1999; 9905-0156563
PR 28 SEP-1999; 9905-0156458
PR 29 SEP-1999; 9905-0156595
PR 04 OCT-1999; 9905-0157117
PR 05 OCT-1999; 9905-0157751
PR 06 OCT-1999; 9905-0157865
PR 07 OCT-1999; 9905-0158079
PR 08 OCT-1999; 9905-0158242
PR 12 OCT-1999; 9905-0158693
PR 13 OCT-1999; 9905-0159294
PR 13 OCT-1999; 9905-0159294
PR 13 OCT-1999; 9905-0159295
PR 14 OCT-1999; 9905-0159429
PR 14 OCT-1999; 9905-0159430
PR 14 OCT-1999; 9905-0159431
PR 14 OCT-1999; 9905-0159647
PR 14 OCT-1999; 9905-0159648
PR 18 OCT-1999; 9905-0159841
PR 21 OCT-1999; 9905-0160741
PR 21 OCT-1999; 9905-0160767
PR 21 OCT-1999; 9905-0160768
PR 21 OCT-1999; 9905-0160770
PR 21 OCT-1999; 9905-0160814
PR 21 OCT-1999; 9905-0160815
PR 22 OCT-1999; 9905-0160980
PR 22 OCT-1999; 9905-0160981
PR 22 OCT-1999; 9905-0160989
PR 25 OCT-1999; 9905-0161404
PR 25 OCT-1999; 9905-0161405
PR 25 OCT-1999; 9905-0161406
PR 26 OCT-1999; 9905-0161359
PR 26 OCT-1999; 9905-0161460
PR 26 OCT-1999; 9905-0161461
PR 26 OCT-1999; 9905-0161920
PR 26 OCT-1999; 9905-0161921
PR 26 OCT-1999; 9905-0161993
PR 26 OCT-1999; 9905-0162142

```

Query Match 5.58; Score 259; DB 21; Length 512;

Host Local Similarity 21.98; Prod. No. 150-15; Index 176; Gaps 21;

Matches 147; Conservative 83; Mismatches 149;

```

27 55 STARTER POLYMERASE SUBUNIT 1106
28 47 ARKININ-1 POLYMERASE SUBUNIT 1106
29 80 HLA-B*08:01 POLYMERASE SUBUNIT 1106
30 107 DYCKIA POLYMERASE SUBUNIT 1106
31 140 MIMIC POLYMERASE SUBUNIT 1106
32 157 VINCENIN POLYMERASE SUBUNIT 1106
33 200 GSK-3B POLYMERASE SUBUNIT 1106
34 214 GSK-3B POLYMERASE SUBUNIT 1106

```

```

260 KRPTELKMKKVFSEVSHLYPTPKASLFLSSLR/ADLTPEQISQ/LKQINDYLAE 319
259 KLTLPETQHWTFQTHPIYLAIS -----PDTVQAKKINE----- 296
320 KSLFVYVIMCIACGLKEKEVKEKILRSKPPICLIPNLEPESGSGGKSSLLD/- 377
297 -----ELVGEVAV -----MQLDNGYLES1 316
378 -----DTVLISLQ/LNKEKDWG/-EAPVPIEDVQSN-----LQHS 414
317 rrrtqrdrvylll ldrtrrpsgylscrtgttsqngpmr lpcaspsvgnwlpah 375
415 NSNNELSAATLPLILKEKTEYQLN-----KILIFQLKATYKKNQIKKENVQI 467
376 vdrgrtgarstpv-----dkwalgshlprcm-ncvkal qelnvewkk----- 424
468 PRIMKCIIMAAIIG -----VGCALHAKYVA--IDKIDP--PDRQIVQIPKCHGYDE 517
425 qhymkewwpqlaqgqlimnqghl rdesstldeccamtspvlkietql ykareeky 484
518 LNSPECHAKFKKRVKAAVYSHNLLVWQGLUSL/ATPL 556
485 lrdlgr -----vnpqfll l d lcaatl 506

```

RESULT 10

AAG36159

ID AAG36159 standard; protein; 452 AA.

XX AAG36159;

AC AAG36159;

XX AAG36159;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 44274.

XX Protein identification: signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KM termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

	Genetic Matrix	5'-3'	Sequence 457-512	100-212	Length 4522
	Test Local Similarity	21-981	Prod. No. 1 to 153	Indels	Gaps
	Matrix	1-21	Conservation	84%	Mismatches 19%
98	25 0871 100992	90085 0161406			
99	26 0871 100992	90085 0161459			
100	26 0871 100992	90085 0161460			
101	26 0871 100992	90085 0161461			
102	26 0871 100992	90085 0161462			
103	26 0871 100992	90085 0161463			
104	26 0871 100992	90085 0161464			
105	26 0871 100992	90085 0161465			
106	26 0871 100992	90085 0161466			
107	26 0871 100992	90085 0161467			
108	26 0871 100992	90085 0161468			
109	26 0871 100992	90085 0161469			
110	26 0871 100992	90085 0161470			
111	26 0871 100992	90085 0161471			
112	26 0871 100992	90085 0161472			
113	26 0871 100992	90085 0161473			
114	26 0871 100992	90085 0161474			
115	26 0871 100992	90085 0161475			
116	26 0871 100992	90085 0161476			
117	26 0871 100992	90085 0161477			
118	26 0871 100992	90085 0161478			
119	26 0871 100992	90085 0161479			
120	26 0871 100992	90085 0161480			
121	26 0871 100992	90085 0161481			
122	26 0871 100992	90085 0161482			
123	26 0871 100992	90085 0161483			
124	26 0871 100992	90085 0161484			
125	26 0871 100992	90085 0161485			
126	26 0871 100992	90085 0161486			
127	26 0871 100992	90085 0161487			
128	26 0871 100992	90085 0161488			
129	26 0871 100992	90085 0161489			
130	26 0871 100992	90085 0161490			
131	26 0871 100992	90085 0161491			
132	26 0871 100992	90085 0161492			
133	26 0871 100992	90085 0161493			
134	26 0871 100992	90085 0161494			
135	26 0871 100992	90085 0161495			
136	26 0871 100992	90085 0161496			
137	26 0871 100992	90085 0161497			
138	26 0871 100992	90085 0161498			
139	26 0871 100992	90085 0161499			
140	26 0871 100992	90085 0161500			
141	26 0871 100992	90085 0161501			
142	26 0871 100992	90085 0161502			
143	26 0871 100992	90085 0161503			
144	26 0871 100992	90085 0161504			
145	26 0871 100992	90085 0161505			
146	26 0871 100992	90085 0161506			
147	26 0871 100992	90085 0161507			
148	26 0871 100992	90085 0161508			
149	26 0871 100992	90085 0161509			
150	26 0871 100992	90085 0161510			
151	26 0871 100992	90085 0161511			
152	26 0871 100992	90085 0161512			
153	26 0871 100992	90085 0161513			
154	26 0871 100992	90085 0161514			
155	26 0871 100992	90085 0161515			
156	26 0871 100992	90085 0161516			
157	26 0871 100992	90085 0161517			
158	26 0871 100992	90085 0161518			
159	26 0871 100992	90085 0161519			
160	26 0871 100992	90085 0161520			
161	26 0871 100992	90085 0161521			
162	26 0871 100992	90085 0161522			
163	26 0871 100992	90085 0161523			
164	26 0871 100992	90085 0161524			
165	26 0871 100992	90085 0161525			
166	26 0871 100992	90085 0161526			
167	26 0871 100992	90085 0161527			
168	26 0871 100992	90085 0161528			
169	26 0871 100992	90085 0161529			
170	26 0871 100992	90085 0161530			
171	26 0871 100992	90085 0161531			
172	26 0871 100992	90085 0161532			
173	26 0871 100992	90085 0161533			
174	26 0871 100992	90085 0161534			
175	26 0871 100992	90085 0161535			
176	26 0871 100992	90085 0161536			
177	26 0871 100992	90085 0161537			
178	26 0871 100992	90085 0161538			
179	26 0871 100992	90085 0161539			
180	26 0871 100992	90085 0161540			
181	26 0871 100992	90085 0161541			
182	26 0871 100992	90085 0161542			
183	26 0871 100992	90085 0161543			
184	26 0871 100992	90085 0161544			
185	26 0871 100992	90085 0161545			
186	26 0871 100992	90085 0161546			
187	26 0871 100992	90085 0161547			
188	26 0871 100992	90085 0161548			
189	26 0871 100992	90085 0161549			
190	26 0871 100992	90085 0161550			
191	26 0871 100992	90085 0161551			
192	26 0871 100992	90085 0161552			
193	26 0871 100992	90085 0161553			
194	26 0871 100992	90085 0161554			
195	26 0871 100992	90085 0161555			
196	26 0871 100992	90085 0161556			
197	26 0871 100992	90085 0161557			
198	26 0871 100992	90085 0161558			
199	26 0871 100992	90085 0161559			
200	26 0871 100992	90085 0161560			

[illegible]

97	482	VGGATLAKTAA	TGKGT	PGRQLEVDHPRGVDELSSVGHAKRRLKA	534
100	664	INUNGNHTDTRDSSSLDDSCAMSPGLVKLRQLQYKRRKKYLIDLP----			432
27	545	WVVSRPDAVVMGRTLSLCAPET	556		
100	433	YNGQFTTL	LD	LCAFL	448

RESULT 12					
XX	AAV1B05.3				
XX	AAV1B05.3 standard: protein, 504 AA.				
XX	AAV1B05.3				
XX	AAV1B05.3				
XX	05_AAV_1999 (first entry)				
XX	766K_20 protein sequence:				
DE					
XX	Transcription unit: MARK2 Kinase; tsx's kinase regulatory region; T66K;				
KM	containing two Responder/Distorter signalling cascades: 1 Responder.				
XX					
XX	Mus sp.				
CS					
F8	W0725415.AZ:				
TX					
10	27 KAV 1999.				
XX					
19	18 NOV 1998: 98WD EP07305				
XX					
19	02 MAR 1998: 98EP 0103396.				
19	18 NOV 1997: 97EP 0120100.				
XX					
1A	(PLAC) MAX PLANCK GES FORSCHUNG WISSENSCHAFTEN.				
11					
E1	Herrmann B., Kisipet A., Koschütz H.				
XX	WFI: 1999 047466/29.				
10	N FSNB: MAX70841.				
XX					
15	Nucleic acids involved in the responder phenotype in mice.				
XX	claim 2: fig 7d: 11/bp: English.				
XX					
15	This sequence represents the T66K_20 protein.				
XX	The invention related to a nucleic acid molecule (1) comprising a				
57	transcription unit encoding in its 5' portion a kinase having a homology				
57	to MARK2 kinase and the 3' portion of the nucleotide sequence has a high				
57	homology to tsx's kinase, a protein produced by transgene creatures				
57	containing (1), are useful for production of offspring, Tsdxk, its				
57	regulatory region, recombinant DNA, vectors, host cells, antibodies,				
57	etc., are useful for the isolation of receptors on the surface of sperm				
57	production of contraceptives, they can also be used to identify chemicals				
57	or biological compounds able to trigger the (promutary) activation or				
57	inhibition of the Responder/Distorter signalling cascade, or to identify				
57	and isolate receptors and other members of the cascade that bind the				
57	transgene animals, and selecting against (1) also provide a means for				
57	distant that the transmission ratio of genetic traits by affecting genes of				
57	the Responder/Distorter signal cascade other than the 1 Responder. They				
57	also allow distortion to a non Mendelian ratio, of the transmission of a				
57	genetic trait, i.e., determination of sex, from male mammals to their				
57	offspring by expressing during spermatogenesis/Spermiogenesis a gene				
57	involved in sperm motility and/or fertilisation. The genes and proteins				
57	involved in the Responder phenotype and Responder/Distorter signalling				
57	cascade, as well as the inventive methods are advantageous in breeding				
57	strategies by allowing for specific selection of genetic traits and in				
57	particular, of sex.				
57	Sequence: 504 AA:				

Query Match	254	5	48	Score	254	DB	20	Length	504
Post Local Similarity	22.2%								
Matches	95	Conservation	76	Mismatches	150	Indels	100	Gaps	15
QY	47	KILDPQLTKTTHPRICQYVDISKGRDLNVVNAHDE	KSLDELLEKKPWSNSTVLC	95					
DB	68	FTSVSEVLIIMADHPHLSLIGYIICK	KKYVIMICKKSIVYGHFKaay/qchcara	126					
QY	96	IAEEVAGLQYMKKHSIVHVALSPINILTEKRGTKLAKFLYMTAHGDEVPFQYVS	155						
DB	127	IKKILISamycchmghvvdrlkpd	lmvckqkvvkldtdlrvkrgqsklntfqrYH	186					
QY	156	YIAEVAAGQIQETTHHMSKKRPPSGKSDVMSAIIIFELICGKRLPQSLDISKLRK	215						
DB	187	LSapevllstpy	gdkpldvwllqavlymrvcklpldacsikr1vkr	243					
QY	216	ELILHVVHPLILVIAEEHGLSDIKELQETVLDLUNKITPHPSKKRPPEHMKDQVSE	275						
DB	244	ILH-----	gkyslprlsaaqslldlntamprlrlpvaamvhpvce	278					
QY	276	VSPFY-----TPTKFSALSSSLRCADLTPEQISQKQKQINNYLAERSLVEVYL	328						
DB	279	qstvtprpccqcpil	krplvavkam-----ghvrtqgdl	328					
QY	429	WGLAGTDLKELVYNKRLKS-----KIPVCLLPPLPRKSSSGDGRDSSSLDPTVY	482						
DB	429	YCL-----IKqdl	krpdprlarpvnsavrlp	466					
QY	484	LSLQVQNKRLKDWGQIAEYPLLDQSNLPSNSNNELSAATLILILKKQILEYQINLI	442						
DB	467	Lq	lrrt-----enqetpwwssarqysvsg	401					
QY	443	ILFDRL	449						
DB	402	vsmpsvl	408						
RESULT	13								
ID	AA885122	standard; Protocol; 326 AA.							
XX	AA885122:								
XX	01-MAR 1996	(first entry)							
DE	CLK6	R31C mutant.							
XX	CD44	cyclical regulatory protein; cyclin-dependent kinase inhibitor							
XX	CD44; CLK6;	cancer; cell proliferation; melanoma.							
XX	Bmo sapiens.								
XX	Key	Location/Qualifications							
FT	Misc-difference 29	/note- "unidentified amino acid"							
FT	Misc-difference 31	/note- "amino acid at position 31 is given as (unidentified/Xaa); but is the site of the Arg to Cys mutation"							
FT	Misc-difference 100	/note- "unidentified amino acid"							
FT	Misc-difference 102	/note- "unidentified amino acid"							
XX	W09528484	AI.							
XX	26-04-1995.								
XX	14 APR 1995.	9500-05046							
XX	29-IV-1994.	9408-0346147.							
XX	14-APR-1994.	9408-0227371.							

```

PR 25-MAY-1994: 9405-024801.2.
FR 14-SEP-1994: 9405-0306511.
XX
XX (COLD-) COLD SPRING HARBOR LAB.
XX
XX Beach DH, Demetrick DJ, Hannon GJ, Serrano M;
XX WPI: 1995-173798/48.
XX
XX New cell cycle regulating proteins bind to cyclin dependent kinase -
XX P1 and related nuclear acids, antihodies etc., used in diagnosis and
XX P1 therapy of abnormal cell proliferation, degeneration etc.
XX
XX Example 10; Page 84-85; 10*pp; English.
XX
XX A mutant (AAR85122) of cyclin-dependent kinase 4 (CDK4) was discovered
XX CC in cells from a melanoma patient. The mutant gene was cloned and
XX CC expressed in Sf9 cells. CDK4 mutant was shown to be insensitive
XX CC to the cell cycle regulatory proteins p16 and p15 (AAR85116-17).
XX CC An Arg31-Cys mutation in CDK6 (AAR85122) is expected to have a
XX CC similar effect. Examination of such mutants will provide a model for
XX CC p16/p15 inhibition of CDK4/CD6.
XX
XX Sequence 126 AA:
XX
XX Query Match 5.4%; Score 253; DB 16; Length 326;
XX Best Local Similarity 28.7%; Pred. No. 2,5e-15;
XX Matches 93; Conservative 48; Mismatches 99; Indels 84; Gaps 16.
XX
XX 25 GSNCLPLPNSIKHSFPHKTHHPRICQYD---ISKGRH-RIVVAVAFHCHSLSD 80
XX I::I::: : : : : : : : : : : : : : : : : : : : : : : :
XX 50 qeeempI--SLTFEAVLTLHLEHPRVVLIDVCTSRIDREKILVLEKXQILL 107
XX
XX R1 LDRRRK---VSGSTVLCATFEVLOSLQYMKHKGIVHRLSPHNLIDRKHGKIKKFL 137
XX 108 yI-dkvepygyprcttkommtqIrguflshrvvhdIppqnlvtssgqkIdfgI 166
XX
XX 138 YHMTAHEDVDVDFPIGPS-----VLAEVTAQGIKFTIDHPSKKRLPSPSKSWSLG 191
XX 167 arI-----yslqmaltsvsvlwlwrafavevllqssyal-----pv-----dlwsvq 206
XX
XX 192 LLEFLCVGKKIKFO-SLDSISERIKELHIDCVDDPLIVLAFPHGCLDI----- 218
XX 207 cIleamrlrplIITGsdvqaykIl-----dYvIapgeedprrvdIprqallksa 259
XX
XX 239 -----IKELPEVIDLNKCLTFEHSKRPEDDELMDKDVSEVSPYTFPKPASFS 292
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 260 qpktektvIdelqkIlIkcltItpakr-----Isaysalshpy-----fgd 302
XX
XX 293 SLRC---ADLTFEDISQCLXDN 313
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 303 IerkenIdshlpp--sqntslsln 324
XX
XX RESULT 14
XX AAY43926
XX ID AAY43926 standard; Protein: 257 AA.
XX
XX AAY43926:
XX
XX 21-DPC-1999 (first entry)
XX
XX Yeast protein kinase #5.
XX
XX Prediction: secondary structure; alignment; evolutionary conservation;
XX KM homology; periodicity; co-variation analysis; antigenic site;
XX site directed mutagenesis; interaction.
XX
XX Saccharomycetes cerevisiae.
XX
XX USS958784-A
XX
XX

```

XX	28-SEP-1999.
XX	
XX	25-MAR-1992: 92US-0857224.
XX	
XX	25-MAR-1992: 92US-0857224.
XX	
XX	(HENN/) HENNER S. A.
XX	
XX	Benner SA:
XX	
XX	WPI: 1999-570766/48.
XX	
XX	Predicting the folded structure of proteins
XX	
XX	Disclosure: Column 185-188; 113pp; English.
XX	
XX	Sequences MAY41902-Y44015 represent proteins used in a novel method of
XX	predicting the folded structure of proteins, by aligning sequences of
XX	homologous proteins and using patterns of evolutionary conservat and
XX	varied sequences to assign positions. Positions in the alignment are
XX	assigned to the surface or inside of the folded structure, active sites,
XX	and parsing segments. Secondary structural units are assigned by
XX	identifying periodicity in the assignments, and assembled into globular
XX	form using distance constraints imposed by disulfide bridges, active
XX	site assignments and co-variation analysis. The predicted secondary
XX	structures are useful for identifying antigenic sites on a protein
XX	molecule, as guides for site directed mutagenesis studies, and for
XX	understanding the interaction of a protein with other molecules.
XX	
XX	Sequence 257 AA:
XX	

[illegible]


```

PR 48-SEP-1999: 990S-0156458.
PR 29-SEP-1999: 990S-0156596.
PR 04-OCT-1999: 990S-0157117.
PR 05-OCT-1999: 990S-0157753.
PR 06-OCT-1999: 990S-0157865.
PR 07-OCT-1999: 990S-0158029.
PR 08-OCT-1999: 990S-0158232.
PR 12-OCT-1999: 990S-0158369.
PR 13-OCT-1999: 990S-0159293.
PR 13-OCT-1999: 990S-0159294.
PR 13-OCT-1999: 990S-0159295.
PR 14-OCT-1999: 990S-0159329.
PR 14-OCT-1999: 990S-0159330.
PR 14-OCT-1999: 990S-0159331.
PR 14-OCT-1999: 990S-0159637.
PR 14-OCT-1999: 990S-0159638.
PR 18-OCT-1999: 990S-0159584.
PR 21-OCT-1999: 990S-0160741.
PR 21-OCT-1999: 990S-0160767.
PR 21-OCT-1999: 990S-0160768.
PR 21-OCT-1999: 990S-0160770.
PR 21-OCT-1999: 990S-0160814.
PR 21-OCT-1999: 990S-0160815.
PR 22-OCT-1999: 990S-0160980.
PR 22-OCT-1999: 990S-0160981.
PR 22-OCT-1999: 990S-0160989.
PR 25-OCT-1999: 990S-0161404.
PR 25-OCT-1999: 990S-0161405.
PR 25-OCT-1999: 990S-0161406.
PR 26-OCT-1999: 990S-0161359.
PR 26-OCT-1999: 990S-0161360.
PR 26-OCT-1999: 990S-0161361.
PR 28-OCT-1999: 990S-0161920.
PR 28-OCT-1999: 990S-0161992.
PR 28-OCT-1999: 990S-0161993.
PR 29-OCT-1999: 990S-0162142.

```

Query Match 5.3%; Score 250.5; DB 21; Length 437;
 Host Locality 24.4%; Pred. No 7.2e-15;
 Matches 108; Conservative 69; Mismatches 187; Indels 79; Gaps 14;

```

QY 36 IKILGRFOLKTIHPRLOQVYDLSRKHEELVVAHCHERSLEDLLREKRPVCSYLC 95
DB 5 IKVQREIKLITIRPHILIRYVELEPRNDILVYVWYVKSQELIDYIVKQKIQEDAEATH 64
QY 96 IAEVLDQIQYNNKKHOLVKAISPNNILDRKCHIKLAKPLVYMTAHQDDVPRIGYPS 155
DB 65 IIGQIISYQVEYCHRMIVARDIKPENVIDSQGLKIVDILQISNMHDPJILIKSCQSPA 124
QY 156 YLAPEYIAGIFRTIDHMSKKPLPSPKSPWLSGILFELCVGRKLEUSDLS---ER 212
DB 125 YAAPEVY-----SAKPY--SPDADVWSEQVILYAILCYLIPIDGEMIPNVTEK 170
QY 213 LKFLITLDCVDDTLIVLAENGLDILKELEPETYIDELINKCLTFIPSKRETPDELMKDRV 272
DB 171 IK-----RAMYLLPHLHSLARDLIPRLMWDPMRLISLELRHPW 212
QY 273 FSEVSPHYTPFKPASPSSSRKADLILPEDIISQLOKDIINN--DYLAERSTREVVYLMC 330
DB 213 Lmhhlpylsi-ppldltlqaklee-elqgnvniqldfnhvdslantlqpealvayh 270
QY 331 IAGGLEKEELVNNKEIRSNP-----ICTLP--NFLFERGESFSGGRD-RSSLDDTT 380
DB 271 IIdnrtngsvprdpfgrskfkolsdqtfnstlpvgnltshvgsfsalygksnvkdkt 330
QY 381 VTLSTCOLRRLKDWGGEAFYPLLEDQSNLP-----HSNSNNE 419
DB 331 wtlqtl-qsgsqpydmteitkal-----gnlkicwkkilqynlkgwvrtstayknhlie 384
QY 420 LSAATPLPILIRKREVOYINRI 442
DB 385 decalllpvlik---leqlykv 404

```

Search completed: June 26, 2001, 03:02:51
 Job time: 1859 Sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2001, 01:23:31 ; Search time 84.37 Seconds

(without alignments)
5889.158 Million cell updates/sec

Title: US-09-707-121-1

Perfect score: 2682
Sequence: 1 atgtctccctgaagagacg.....tccatctctcaataatga 2682

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapexl 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgm2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgm2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgm2_6/ptodata/2/ina/5A_COMB.seq:*
- 4: /cgm2_6/ptodata/2/ina/5B_COMB.seq:*
- 5: /cgm2_6/ptodata/2/ina/PTCUS_COMB.seq:*
- 6: /cgm2_6/ptodata/2/ina/backlitcsl.seq:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57.2	2.1	7218	1 US-08-232-463-14	Sequence 14, Appl
2	50.6	1.9	1647	4 US-09-101-146-44	Sequence 44, Appl
3	47	1.8	2754	4 US-09-429-322-3	Sequence 3, Appl
4	42.2	1.6	4363	2 US-08-685-576-5	Sequence 5, Appl
5	40.6	1.5	5053	2 US-08-685-576-2	Sequence 2, Appl
6	40.4	1.5	3018	2 US-08-860-150-6	Sequence 6, Appl
7	40.4	1.5	3018	2 US-08-338-132-6	Sequence 6, Appl
8	39.8	1.5	2104	4 US-09-313-930-1	Sequence 1, Appl
9	39.4	1.5	1541	6 5183884-1	Patent No. 5183884
10	39.4	1.5	1542	1 US-07-978-895-1	Sequence 1, Appl
11	39.4	1.5	1542	1 US-08-473-119-1	Sequence 1, Appl
12	39.4	1.5	1542	1 US-08-475-152-1	Sequence 1, Appl
13	38.6	1.4	2500	5 PCT-US95-05008-3	Sequence 3, Appl
14	38.2	1.4	1251	2 US-09-211-930-2	Sequence 2, Appl
15	38.2	1.4	1251	2 US-09-340-993-2	Sequence 2, Appl
16	38.2	1.4	1353	2 US-09-211-930-8	Sequence 8, Appl
17	38.2	1.4	1353	2 US-09-340-993-8	Sequence 8, Appl
18	38.2	1.4	3201	2 US-09-211-930-1	Sequence 1, Appl
19	38.2	1.4	3201	2 US-09-340-993-1	Sequence 1, Appl
20	37.8	1.4	903	2 US-08-874-147-9	Sequence 9, Appl
21	37.8	1.4	903	3 US-08-093-522-9	Sequence 9, Appl
22	37.6	1.4	2747	2 US-08-874-147-1	Sequence 1, Appl
23	37.6	1.4	2747	2 US-09-093-522-1	Sequence 1, Appl
24	37	1.4	2268	1 US-08-444-005-14	Sequence 14, Appl
25	37	1.4	4112	4 US-09-060-410-1	Sequence 1, Appl
26	37	1.4	4545	6 5183884-3	Patent No. 5183884
27	37	1.4	4905	1 US-07-978-895-3	Sequence 3, Appl

28	37	1.4	4905	1 US-08-473-119-3	Sequence 3, Appl
29	37	1.4	4905	2 US-08-475-352-3	Sequence 3, Appl
30	36.4	1.4	1448	2 US-08-042-218A-1	Sequence 1, Appl
31	36.4	1.4	1881	1 US-08-610-731A-1	Sequence 1, Appl
32	36.4	1.4	1881	2 US-09-067-379-1	Sequence 1, Appl
33	36.4	1.4	1881	3 US-09-067-506-1	Sequence 1, Appl
34	36.4	1.4	2244	4 US-09-094-714A-48	Sequence 48, Appl
35	36.2	1.3	1544	2 US-08-837-593-1	Sequence 1, Appl
36	35.8	1.3	1424	4 US-08-403-634-3	Sequence 3, Appl
37	35.8	1.3	1424	4 US-08-013-441B-1	Sequence 3, Appl
38	35.8	1.3	2434	2 US-08-540-804-15	Sequence 15, Appl
39	35.8	1.3	2434	2 US-08-218-265-15	Sequence 15, Appl
40	35.8	1.3	2434	3 US-08-521-872-15	Sequence 15, Appl
41	35.8	1.3	2434	4 US-08-590-399-15	Sequence 15, Appl
42	35.8	1.3	4739	3 US-08-685-871-1	Sequence 1, Appl
43	35.6	1.3	2248	1 US-08-369-780-1	Sequence 1, Appl
44	35.6	1.3	2248	1 US-08-475-682-1	Sequence 1, Appl
45	35.6	1.3	2248	1 US-08-780-833-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14

? Sequence 14, Application US/08232463

? Patent No. 5670367

? GENERAL INFORMATION:

? APPLICANT: DORNER, F.

? APPLICANT: SCHEFFLINER, F.

? TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

? NUMBER OF SEQUENCES: 52

? CORRESPONDENCE ADDRESS:

? ADDRESSEE: Foley & Lardner

? STREET: 1800 Diagonal Road, Suite 500

? CITY: Alexandria

? STATE: VA

? COUNTRY: USA

? ZIP: 22313-0299

? COMPUTER READABLE FORM:

? MEDIUM TYPE: Floppy disk

? COMPUTER: IBM PC compatible

? OPERATING SYSTEM: PC-DOS/MS-DOS

? SOFTWARE: Patent in Release #1.0, Version #1.25

? CURRENT APPLICATION DATA:

? APPLICATION NUMBER: US/08/232,463

? FILING DATE:

? CLASSIFICATION: 435

? PRIOR APPLICATION DATA:

? APPLICATION NUMBER: US/07/935,313

? FILING DATE:

? APPLICATION NUMBER: EP 91 114 300.6

? FILING DATE: 26-AUG-1991

? ATTORNEY/AGENT INFORMATION:

? NAME: BENT, Stephen A.

? REGISTRATION NUMBER: 29,768

? REFERENCE/DOCKET NUMBER: 30472/114 IMM

? TELECOMMUNICATION INFORMATION:

? TELEPHONE: (703)836-9300

? TELEFAX: (703)683-4109

? TELEFAX: 899149

? INFORMATION FOR SEQ ID NO: 14:

? SEQUENCE CHARACTERISTICS:

? LENGTH: 7218 base pairs

? TYPE: nucleic acid

? STRANDEDNESS: single

? TOPOLOGY: linear

? IMMEDIATE SOURCE:

? CLONE: PTZapl-F15

? US-08-232-463-14

Mathematics: 60; Computer and Information Sciences: 41; Index: 0; Maps: 0

APPLICANT: KLAUS, MATTHIAS H.; AARONSON, STUART A

TELECOMMUNICATION INFORMATION

TELEPHONE: (404) 688-0770

TELEFAX: (404) 688-9880

INFORMATION FOR SEQ ID NO: 1

SEQUENCE CHARACTERISTICS

LENGTH: 1542 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

FEATURE

NAME/KEY: exon

LOCATION: 66..221

NAME/KEY: exon

LOCATION: 780..855

NAME/KEY: exon

LOCATION: 1040..1185

NAME/KEY: 10100

LOCATION: 222..779

NAME/KEY: 10100

LOCATION: 856..1039

NAME/KEY: CDS

LOCATION: join(66..221, 780..855, 1040..1185)

US-08-475-119-1

QUALITY MATCH

Host local similarity: 55.5%; Pred. No. 0.075;

Matches: 76; Conservative: 0; Mismatches: 61; Indels: 0; Gaps: 0;

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

FILING DATE:

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 07/978,895

FILING DATE:

APPLICATION NUMBER: US 07/444,406

FILING DATE: 01-DEC-1989

ATTORNEY/AGENT INFORMATION:

NAME: Perryman, David G.

REGISTRATION NUMBER: 31,438

REFERENCE/DOCKET NUMBER: 1414-028

TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 688-9880

TELEFAX: (404) 688-0770

INFORMATION FOR SEQ ID NO: 1

SEQUENCE CHARACTERISTICS:

LENGTH: 1542 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

FEATURE

NAME/KEY: exon

LOCATION: 66..221

NAME/KEY: exon

LOCATION: 780..855

NAME/KEY: exon

LOCATION: 1040..1185

NAME/KEY: 10100

LOCATION: 222..779

NAME/KEY: 10100

LOCATION: 856..1039

NAME/KEY: CDS

LOCATION: join(66..221, 780..855, 1040..1185)

US-08-475-352-1

QUALITY MATCH

Host local similarity: 1.5%; Score 39.4; DB 2;

Matches: 76; Conservative: 0; Mismatches: 61; Indels: 0; Gaps: 0;

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

FILING DATE:

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 07/978,895

FILING DATE:

APPLICATION NUMBER: US 07/444,406

FILING DATE: 01-DEC-1989

ATTORNEY/AGENT INFORMATION:

NAME: Perryman, David G.

REGISTRATION NUMBER: 31,438

REFERENCE/DOCKET NUMBER: 1414-028

TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 688-9880

TELEFAX: (404) 688-0770

INFORMATION FOR SEQ ID NO: 1

SEQUENCE CHARACTERISTICS:

LENGTH: 1542 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

FEATURE

NAME/KEY: exon

LOCATION: 66..221

NAME/KEY: exon

LOCATION: 780..855

NAME/KEY: exon

LOCATION: 1040..1185

NAME/KEY: 10100

LOCATION: 222..779

NAME/KEY: 10100

LOCATION: 856..1039

NAME/KEY: CDS

LOCATION: join(66..221, 780..855, 1040..1185)

US-08-475-352-1

QUALITY MATCH

Host local similarity: 1.5%; Score 39.4; DB 2;

Matches: 76; Conservative: 0; Mismatches: 61; Indels: 0; Gaps: 0;

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

FILING DATE:

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 07/978,895

FILING DATE:

APPLICATION NUMBER: US 07/444,406

FILING DATE: 01-DEC-1989

ATTORNEY/AGENT INFORMATION:

NAME: Perryman, David G.

REGISTRATION NUMBER: 31,438

REFERENCE/DOCKET NUMBER: 1414-028

TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 688-9880

TELEFAX: (404) 688-0770

INFORMATION FOR SEQ ID NO: 1

SEQUENCE CHARACTERISTICS:

LENGTH: 1542 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

FEATURE

NAME/KEY: exon

LOCATION: 66..221

NAME/KEY: exon

LOCATION: 780..855

NAME/KEY: exon

LOCATION: 1040..1185

NAME/KEY: 10100

LOCATION: 222..779

NAME/KEY: 10100

LOCATION: 856..1039

NAME/KEY: CDS

LOCATION: join(66..221, 780..855, 1040..1185)

US-08-475-352-1

QUALITY MATCH

Host local similarity: 1.5%; Score 39.4; DB 2;

Matches: 76; Conservative: 0; Mismatches: 61; Indels: 0; Gaps: 0;

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

FILING DATE:

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 07/978,895

FILING DATE:

APPLICATION NUMBER: US 07/444,406

FILING DATE: 01-DEC-1989

ATTORNEY/AGENT INFORMATION:

NAME: Perryman, David G.

REGISTRATION NUMBER: 31,438

REFERENCE/DOCKET NUMBER: 1414-028

TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 688-9880

TELEFAX: (404) 688-0770

INFORMATION FOR SEQ ID NO: 1

SEQUENCE CHARACTERISTICS:

LENGTH: 1542 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

FEATURE

NAME/KEY: exon

LOCATION: 66..221

NAME/KEY: exon

LOCATION: 780..855

NAME/KEY: exon

LOCATION: 1040..1185

NAME/KEY: 10100

LOCATION: 222..779

NAME/KEY: 10100

LOCATION: 856..1039

NAME/KEY: CDS

LOCATION: join(66..221, 780..855, 1040..1185)

US-08-475-352-1

QUALITY MATCH

Host local similarity: 1.5%; Score 39.4; DB 2;

Matches: 76; Conservative: 0; Mismatches: 61; Indels: 0; Gaps: 0;

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

DB

FILING DATE:

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 07/978,895

FILING DATE:

APPLICATION NUMBER: US 07/444,406

FILING DATE: 01-DEC-1989

ATTORNEY/AGENT INFORMATION:

NAME: Perryman, David G.

REGISTRATION NUMBER: 31,438

REFERENCE/DOCKET NUMBER: 1414-028

TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 688-9880

TELEFAX: (404) 688-0770

INFORMATION FOR SEQ ID NO: 1

SEQUENCE CHARACTERISTICS:

LENGTH: 1542 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

